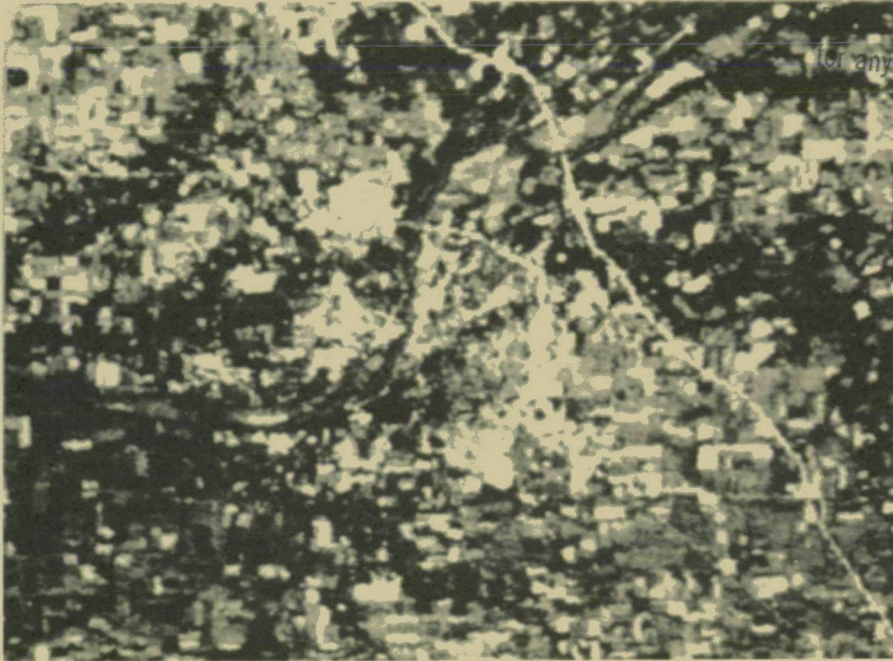


A STUDY OF THE UTILIZATION OF
ERTS-1 DATA FROM THE WABASH RIVER BASIN

FINAL REPORT FOR NASA CONTRACT NAS5-21773



"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
information and without liability
for any use made thereof."



DAVID A. LANDGREBE & STAFF
PURDUE UNIVERSITY
LABORATORY FOR APPLICATIONS OF REMOTE SENSING
1220 POTTER DRIVE
WEST LAFAYETTE, INDIANA

AUGUST 1974
TYPE III REPORT FOR PERIOD JULY 1, 1972-MAY 31, 1974

PREPARED FOR:
GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND 20331

Original photography may be purchased from
EROS Data Center
10th and Dakota Avenue
Sioux Falls, SD 57198

Cover Photo: ERTS-1 Data collected October 19, 1972 (Band 5, .6 - .7 μ m) showing the Lafayette, Indiana area and Wabash Valley. The Wabash River enters left center and curves north exiting top right.

| | | | | | |
|--|--|--------------------------------------|--|---|--|
| 1. Report No. | | 2. Government Accession No. | | 3. Recipient's Catalog No. | |
| 4. Title and Subtitle A STUDY OF THE UTILIZATION OF ERTS-1 DATA FROM THE WABASH RIVER BASIN; FINAL REPORT | | | | 5. Report Date October 1974 | |
| | | | | 6. Performing Organization Code | |
| 7. Author(s) D.A. Landgrebe & Staff of Purdue Lab for Applications of Remote Sensing | | | | 8. Performing Organization Report No. | |
| 9. Performing Organization Name and Address Purdue University, Laboratory for Applications of Remote Sensing 1220 Potter Drive West Lafayette, Indiana 47906 | | | | 10. Work Unit No. | |
| | | | | 11. Contract or Grant No. NAS5-21773 | |
| 12. Sponsoring Agency Name and Address NASA Goddard Space Flight Center Greenbelt Road Greenbelt, Maryland 20771 | | | | 13. Type of Report and Period Covered Type III July 1, 1972 May 31, 1974 | |
| | | | | 14. Sponsoring Agency Code | |
| 15. Supplementary Notes | | | | | |
| 16. Abstract Nine projects were defined for the study; five ERTS data applications experiments and four supporting technology tasks. 1., The Identification and Area Estimation of Crops experiment tested the usefulness of ERTS data for crop survey and produced results indicating that crop statistics could be obtained from ERTS data. 2., Soil Associations Mapping. Results showed that strong relationships exist between ERTS data derived maps and conventional soil maps. 3., Urban Land Use Analysis experiment results indicate potential for accurate gross land use mapping from ERTS data. 4., Water Resources mapping demonstrated the feasibility of mapping water bodies from ERTS data. 5., Earth Surface Features Identification demonstrated feasibility for deriving large area land use data from ERTS. The four supporting technology tasks provided support for the applications projects through analysis technique development, reformatting and temporal overlay, atmospheric modeling, and scene-processed ERTS data evaluation. | | | | | |
| 17. Key Words (Selected by Author(s)) Digital ERTS-1 Imagery Analysis, Earth Surface Feature Classification | | | | 18. Distribution Statement | |
| 19. Security Classif. (of this report) | | 20. Security Classif. (of this page) | | 21. No. of Pages 353 | |
| | | | | 22. Price* | |

*For sale by the Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia 22151.

iii PREFACE

A. Objectives

The objectives of the work reported herein are outlined in the Data Analysis Plan for the ERTS Investigation "A study of the Utilization of ERTS-1 Data from the Wabash River Basin", ERTS-1 proposal No. SR049. The general objectives are: (1) to evaluate the applications of ERTS-1 measurements which have been appropriately reduced for use in specific earth resources problems, and (2) to determine the desirable measurements needed in future earth resources systems.

B. Scope of Work

There are five scientific investigations which were pursued to evaluate the applications of ERTS-1 measurements to specific earth resources problems. To further support these objectives four specific supporting technology tasks are also included. The nine tasks are all based on the use of digital computer techniques, including the LARSYS multi-spectral analysis system, for studying ERTS data in digital form.

C. Conclusions

The conclusions for this study are numerous and are included in each of the nine sections describing results for each project. The overall conclusions are that ERTS-1 data shows potential for crop mapping, urban land use mapping, and water resources mapping. Less favorable results were observed for soil association mapping and earth surface features identification; however these studies were limited in scope.

D. Summary of Recommendations

The project recommendation came from the Crop Classifications project. First, it was recommended that the number of spectral bands be increased to hopefully enable increased classification accuracy. Also, the spatial resolution was judged too coarse for small agricultural fields and a somewhat smaller IFOV would be recommended for future systems. Thirdly, an increase in frequency of coverage should be considered to enable classification at optimum times and to improve the utilization of the temporal dimension. Lastly, the time lag between the observation and receipt of imagery was judged to be excessive and faster generation of quick look images was recommended.

iv Table of Contents

| | <u>Page</u> |
|---|-------------|
| 1.0 Introduction | 1 |
| 2.0 Identification and Area Estimation of Agricultural Crops by Computer Classification of ERTS-1 MSS Data | 3 |
| 2.1 Introduction | 3 |
| 2.2 Objectives | 4 |
| 2.3 Illinois Study | 4 |
| 2.31 Objectives | 6 |
| 2.32 Procedures | 6 |
| 2.33 Results and Discussion | 10 |
| 2.331 Classification of Test Fields | 10 |
| 2.332 Use of a priori Probabilities in Classifications | 12 |
| 2.333 Unbiasing Classification Results | 17 |
| 2.334 Acreage Estimation | 18 |
| 2.335 Extendability of Training Statistics | 22 |
| 2.336 Number of Training Fields | 24 |
| 2.337 Variability of Classification Results | 26 |
| 2.338 Wavelength Band Selection | 28 |
| 2.339 Spectral Characteristics | 30 |
| 2.340 Utilization of the Spatial Dimension of ERTS Data | 32 |
| 2.341 Utilization of the Temporal Dimension of ERTS Data | 33 |
| 2.342 Spatial Resolution and Field Size | 42 |
| 2.34 Summary and Conclusions | 44 |
| 2.4 Indiana Study | 44 |
| 2.41 Objectives | 44 |
| 2.42 Procedures | 45 |
| 2.43 Results and Discussion | 51 |
| 2.431 Classification of Test Fields | 51 |
| 2.432 Acreage Estimates | 52 |
| 2.433 Summary and Conclusions | 57 |
| 2.5 Missouri and Idaho Studies | 57 |
| 2.51 Analysis of Missouri Data | 57 |
| 2.52 Analysis of Idaho Data | 60 |
| 2.53 Summary and Conclusions | 60 |
| 2.6 Conclusions from Crop Identification Studies | 62 |
| 2.7 References for Section 2.0 | 66 |

| | <u>Page</u> |
|--|-------------|
| 3.0 Soil Association Mapping | 67 |
| 3.1 Introduction | 67 |
| 3.2 White County, Indiana Analysis | 68 |
| 3.21 Soil Association Discussion | 68 |
| 3.22 Use of ERTS Imagery in Field Mapping | 72 |
| 3.23 Conclusions | 74 |
| 3.3 Tippecanoe County, Indiana | 74 |
| 3.31 Procedures | 74 |
| 3.32 Results | 78 |
| 3.33 Spectral Composition of Soil Associations | 80 |
| 3.34 Conclusions | 82 |
| 3.4 Finney County, Kansas | 83 |
| 3.41 Conclusions | 86 |
| 3.5 Summary and Conclusions | 86 |
| 4.0 Urban Land Use Analysis | 90 |
| 4.1 Introduction | 90 |
| 4.11 Scope of Report | 92 |
| 4.2 Milwaukee County Subframe Analysis | 92 |
| 4.21 Data Processing | 92 |
| 4.22 Areal Distribution of Classes | 93 |
| 4.221 Explanation of Classification Image | 93 |
| 4.222 Location and Characteristics of Spectral Classes | 95 |
| 4.3 Chicago Subframe Analysis | 99 |
| 4.31 Data Processing | 99 |
| 4.32 Areal Distribution of Classes | 99 |
| | 100 |
| 4.4 Marion County Subframe Analysis | |
| 4.41 Introduction | 100 |
| 4.411 Similarity between the Milwaukee and Indianapolis Data | 101 |
| 4.412 Differences between the Milwaukee and Indianapolis Subframes | 101 |
| 4.42 Data Processing | 102 |
| 4.43 Classification Results | 102 |

| | <u>Page</u> |
|---|-------------|
| 4.431 Explanation of the Classification Image | 102 |
| 4.432 Areal Distribution and Characteristics of Spectral Classes | 103 |
| 4.44 Accuracy of Classification | 105 |
| 4.5 Further Analysis of Milwaukee and Chicago Subframes | 110 |
| 4.51 Introduction | 110 |
| 4.52 Data Analysis | 110 |
| 4.53 Reclassification of Chicago Subframe | 113 |
| 4.6 Use of Geometrically Corrected ERTS Data | 113 |
| 4.61 Discussion | 113 |
| 4.62 Geometrically Corrected ERTS Classification Results | 115 |
| 4.7 Use of Differing Histograms in Conjunction with the LARS Digital Image Display | 118 |
| 4.71 Discussion | 118 |
| 4.72 "Urban" versus "Rural" Histograms: The Example of the Milwaukee Subframe | 118 |
| 4.8 Analysis of Gary, Indiana Area Subframe | 121 |
| 4.81 Data Processing | 121 |
| 4.82 Classification Results | 121 |
| 4.83 Industrial Land Use Classification | 127 |
| 4.84 Classification Accuracy | 131 |
| 4.85 Area Calculations | 133 |
| 4.9 Conclusions | 133 |
| 5.0 Water Resources Research | 134 |
| 5.1 Introduction | 134 |
| 5.11 Objectives | 134 |
| 5.2 Multispectral Classification of Lakes Freeman and Shafer from ERTS-1 MSS Data | 135 |
| 5.3 Multispectral Classification of Lake Freeman from Aircraft MSS Data | 136 |
| 5.4 Discussion of Satellite and Aircraft Results | 142 |
| 5.5 Multispectral Classification of the Salamonie Reservoir from ERTS-1 Data | 143 |
| 5.6 Water Acreage Estimation from ERTS-1 MSS Data | 146 |
| 5.7 Conclusions and Recommendations | 161 |

| | <u>Page</u> |
|--|-------------|
| 6.0 Earth Surface Features Identification | 163 |
| 6.1 Introduction | 163 |
| 6.2 Background | 164 |
| 6.3 Goals and Objectives | 165 |
| 6.4 Approach | 166 |
| 6.41 Cooperating Agencies | 166 |
| 6.42 Study Area | 167 |
| 6.43 Data Identification and Storage | 167 |
| 6.5 Procedure | 170 |
| 6.6 Results | 173 |
| 6.7 Conclusions | 187 |
| 7.0 Analysis Technique Development | 188 |
| 7.1 Introduction | 188 |
| 7.2 Data-Based Criterion for Defining Training Classes | 188 |
| 7.3 Adaptive Classification | 194 |
| 7.4 Use of Context | 200 |
| 7.5 Layered Classifiers | 212 |
| 7.51 Design for Maximal Accuracy | 214 |
| 7.52 The Search Approach | 223 |
| 7.53 The Search Procedure | 224 |
| 7.54 The Evaluation Function | 229 |
| 7.55 Summary | 234 |
| 7.6 Analysis Technique Development: Concluding Remarks | 239 |
| 8.0 Data Reformatting and Temporal Overlay | 240 |
| 8.1 Introduction | 240 |
| 8.2 Data Reformatting | 242 |
| 8.21 Planned Data Flow | 242 |
| 8.22 Actual Data Flow | 245 |
| 8.23 Reformatting Procedure | 246 |
| 8.24 Reformatting Software | 247 |
| 8.25 Data Products | 256 |
| 8.3 Temporal Registration | 259 |
| 8.4 Data Quality Evaluation | 266 |
| 8.5 Geometric Correction | 271 |
| 8.51 MSS Digital Data Geometric Characteristics | 272 |
| 8.52 Geometric Correction Algorithm | 274 |
| 8.53 Intersample Interpolation | 284 |
| 8.54 Geometric Correction Using Ground Control Points | 288 |
| 8.6 Conclusions | 296 |

| | <u>Page</u> |
|--|-------------|
| 9.0 Atmospheric Correction of Remotely Sensed Multispectral Data | 292 |
| 9.1 Introduction | 292 |
| 9.2 Model Development | 292 |
| 9.3 Model Evaluation | 293 |
| 9.4 Parameter Specification | 300 |
| 9.5 Application Example | 306 |
| 10.0 Comparison of System Corrected and Scene Corrected CCT Data | 311 |
| 10.1 Introduction | 311 |
| 10.2 Data Product Description | 311 |
| 10.3 Description of Scene Corrected Data | 312 |
| 10.4 The Quality Study | 313 |
| 10.5 Conclusions | 323 |
| 11.0 Conclusion of Study and Recommendations | 324 |
| 12.0 Acknowledgements | 331 |
| 13.0 Publications Resulting From Study | 332 |
| 14.0 Image Descriptor Forms | 334 |

v List Of Figures

| <u>Figures</u> | <u>Content</u> | <u>Page</u> |
|----------------|--|-------------|
| | <u>Identification of Crops</u> | |
| 2.1 | ERTS Imagery of Northern Illinois | 2 |
| 2.2 | Computer Classification of Northern Illinois | 11 |
| 2.3 | Distribution of Test Field Performance | 13 |
| 2.4 | Training Set Extendability | 23 |
| 2.5 | Influence of Training Set Size | 25 |
| 2.6 | Mean Relative Spectral Response of Corn and Soybeans for three dates | 38 |
| 2.7 | Mean ERTS Spectral Response for two Channels and three dates | 39 |
| 2.8 | Indiana Map | 46 |
| 2.9 | ERTS Imagery of Indiana | 47 |
| | <u>Soil Association Mapping</u> | |
| 3.1 | Black and White Reproduction of Simulated Color IR from ERTS Data - Soil Map Overlay | 69 |
| 3.2 | Black and White Reproduction of Simulated Color IR from ERTS Data - Soil Map Overlay | 73 |
| 3.3 | Simulated Color IR Reproduction of ERTS Data - Tippecanoe County, Indiana | 75 |
| 3.4 | Classification Result from Tippecanoe County Data | 77 |
| 3.5 | Multispectral Signatures of Soil Classes | 79 |
| 3.6 | ERTS Image from Finney County, Kansas | 84 |
| | <u>Urban Land Use Analysis</u> | |
| 4.1 | Land Use Classification Results for Milwaukee | 94 |
| 4.2 | Land Use Classification Results for Cook County, Illinois | 98 |
| 4.3 | Land Use Classification Results for Marion County, Indiana | 111 |
| 4.4 | Second Iteration Classification of Milwaukee County, Wisconsin | 111 |
| 4.5 | Second Iteration Classification of Cook County, Illinois | 114 |
| 4.6 | Milwaukee County Classification Geometrically Corrected | 114 |

| <u>Figure</u> | <u>Content</u> | <u>Page</u> |
|--|---|-------------|
| 4.7 | Computer Line Printer Classification of Milwaukee County | 116 |
| 4.8 | Marion County Computer Line Printer Classification | 117 |
| 4.9 | Grayscale Images of Milwaukee County area | 119 |
| 4.10 | Grayscale Image of Gary, Indiana area | 122 |
| 4.11 | Comparison of Land Use Classifications and Grayscale Images | 123 |
| 4.12 | Classification Results for Gary, Indiana area | 124 |
| 4.13 | Gary Classification - Northern Half | 128 |
| <u>Water Resources Research</u> | | |
| 5.1 | Spectral Plot of Seven Classes of Water | 137 |
| 5.2 | Printout of Seven Classes of Water in Lake Freeman, Indiana | 139 |
| 5.3 | Aircraft Scanner Imagery of Lake Freeman | 140 |
| 5.4 | Spectral Response Graph Across Lake Freeman | 141 |
| 5.5 | Multispectral Classification of Salamonie Reservoir | 144 |
| 5.6 | Relative Spectral Response of Salamonie Reservoir | 147 |
| 5.7 | Classification Results for Lake Freeman | 152 |
| 5.8 | Lake Edge Model | 154 |
| 5.9 | Percent Correction for ERTS Acreages | 159 |
| 5.10 | Estimated Acreages vs. Actual Acreages | 160 |
| <u>Earth Surface Features Identification</u> | | |
| 6.1 | East Central Tippecanoe County Map | 168 |
| 6.2 | Existing Forest Cover in Data Bank | 174 |
| 6.3 | Forest Cover in Test Site Extracted from Photography | 175 |
| 6.4 | Forest Cover Extracted from ERTS Data Classification | 176 |
| 6.5 | Comparison of 6.3 and 6.4 | 177 |
| 6.6 | Existing Cover - Extracted at 1/10 km Square | 181 |
| 6.7 | Forest Cvoer Extracted from ERTS at 1/10 km Square | 182 |
| 6.8 | Comparison Between Existing and ERTS Data at 1/10 km Square | 183 |

| <u>Figure</u> | <u>Content</u> | <u>Page</u> |
|---------------|--|-------------|
| 6.9 | Existing Forest Cover - Extracted at 1/5 km square | 184 |
| 6.10 | Forest Cover Extracted from ERTS at 1/5 km square | 185 |
| 6.11 | Comparison Between Existing and ERTS Data at 1/5 km square | 186 |

Analysis Technique Development

| | | |
|------|--|-----|
| 7.1 | Selection of Updating Vectors | 197 |
| 7.2 | Basic RIMPAR Flow Chart | 206 |
| 7.3 | Per-Point Clustered Satellite Image | 210 |
| 7.4 | Clustered Partitioned Image | 210 |
| 7.5 | A Binary Decision Tree for Four Class Classification | 216 |
| 7.6 | Flow Chart of the Binary Decision Tree | 218 |
| 7.7 | Classification Results Comparison | 221 |
| 7.8 | Classification Results Comparison | 222 |
| 7.9 | Another Binary Decision Tree for Four Class Classification | 225 |
| 7.10 | Different Results Due to Different Tree Structures | 225 |
| 7.11 | A Tree Structure with Its String | 226 |
| 7.12 | Flow Chart of Search Procedure | 228 |
| 7.13 | Performance of Decision Tree Classifiers on Real and Simulated Data Sets | 232 |
| 7.14 | Predicted and measured Computation Time | 233 |
| 7.15 | Performance of Decision Tree Classifiers | 235 |
| 7.16 | A Tree Structure for an ERTS Example | 237 |

Data Reformatting

| | | |
|-----|---|-----|
| 8.1 | ERTS Data Flow Diagram | 243 |
| 8.2 | LARS ERTS Data Descriptor Form | 248 |
| 8.3 | Example Statistics Program Output | 250 |
| 8.4 | One Line ERTS File Listing Example | 253 |
| 8.5 | Test Correlation Example | 263 |
| 8.6 | Difference Images for Registration Evaluation | 264 |
| 8.7 | Example of Striping Distortion | 267 |
| 8.8 | Geometrically Corrected Data | 283 |
| 8.9 | Nearest Neighbor Error Diagram | 285 |

| <u>Figure</u> | <u>Content</u> | <u>Page</u> |
|--|---|-------------|
| <u>Atmospheric Correction</u> | | |
| 9.1 | Model Block Diagram | 296 |
| 9.2 | Intensity of Scattered Radiation for two Values of Surface Reflectivity | 298 |
| 9.3 | Intensity versus ΔX and r_{\max} | 299 |
| 9.4 | Average ERTS MSS Filter Response Curves | 303 |
| 9.5 | Variation in Upward Radiance with Solar Zenith Angle | 305 |
| 9.6 | Upward Radiance in Band 6 | 307 |
| 9.7 | Upward Radiance in Band 7 | 308 |
| <u>Comparison of System Corrected and Scene Corrected Data</u> | | |
| 10.1 | Image of Border Area Data of Scene Corrected CCT Data of Frame E-1071-16111-601 | 314 |
| 10.2 | Scene Corrected Image Showing Lateral Discontinuity | 314 |
| 10.3 | Horizontal Discontinuity Example in Scene Corrected Data Frame 1071-16111-601 | 315 |
| 10.4 | Scene Corrected Data Showing Horizontal and Vertical Discontinuity | 315 |
| 10.5 | Scene Corrected Data Showing Horizontal and Vertical | 316 |
| 10.6 | Horizontal and Vertical Discontinuity Examples in Scene Corrected Data | 316 |
| 10.7 | System Corrected CCT Data from Same Area as Figures 10.4 thru 10.6 (Band 4) | 318 |
| 10.8 | Same as 10.7 for Band 5 | 318 |
| 10.9 | Same as 10.7 for Band 6 | 319 |
| 10.10 | Scene Corrected Data Showing Test Areas in Water | 319 |
| 10.11 | Test Blocks from Scene Corrected Data | 320 |
| 10.12 | Test Blocks in Scene Corrected Data over Mississippi River | 320 |

vi List of Tables

| | <u>Identification of Crops</u> | <u>Page</u> |
|------|---|-------------|
| 2.1 | Estimated Land Use - Northern Illinois | 7 |
| 2.2 | USDA Estimates of Land Use | 8 |
| 2.3 | Classification Results - DeKalb County | 14 |
| 2.4 | Classification Results - Ogle County | 14 |
| 2.5 | Classification Results - Lee County | 14 |
| 2.6 | Classification Results - with the without prior Probabilities | 16 |
| 2.7 | Classification Results - Equal and Unequal weights | 19 |
| 2.8 | Comparison of USDA Acreages and Classification Averages | 21 |
| 2.9 | Variation in Test Field Classification | 27 |
| 2.10 | Interclass Divergence for ERTS Bands | 29 |
| 2.11 | Mean and Standard Deviation of Spectral Responses | 31 |
| 2.12 | Comparison of Classification Performance Sample versus Point Classification | 34 |
| 2.13 | Comparison of Training and Test Classification for Three Dates | 36 |
| 2.14 | Mean Spectral Response for Three Dates | 36 |
| 2.15 | Comparison of Classification Results Using Spectral and Temporal Dimensions | 41 |
| 2.16 | Summary of Land Use in Northwestern Indiana | 48 |
| 2.17 | Estimated Crop Acreages in Northwestern Indiana | 49 |
| 2.18 | Training and Test Field Classification Performance | 50 |
| 2.19 | Comparison of Test Field Classification Results | 53 |
| 2.20 | ERTS Classification Results - 12 Counties Local and Non-Local Training | 54 |
| 2.21 | Comparison of USDA and ERTS Acreage Estimates | 55 |
| 2.22 | Training Field Classification Results for 29 Segments | 59 |
| 2.23 | Comparison of Spectral and Spectral/Temporal Classification Results for Cotton and Soybeans | 59 |
| 2.24 | Classification Performance of Test Fields - Idaho | 61 |

| | <u>Urban Land Use Analysis</u> | <u>Page</u> |
|-----|---|-------------|
| 4.1 | Classification Accuracy for Marion County | 106 |
| 4.2 | Statistics for ERTS Data from Marion County | 109 |
| 4.3 | Features of Interest in Gary, Indiana Classification | 126 |
| 4.4 | Classification Accuracy for Gary, Indiana Classification | 129 |
| 4.5 | Land Use Area Calculations for Gary Area | 132 |
| | <u>Water Resources Research</u> | |
| 5.1 | Mean Response of Spectral Classes of Water in Salamonie Reservoir | 145 |
| 5.2 | Surface Area of Lakes in Indiana | 149 |
| 5.3 | Initial Water Acreages from ERTS Classification | 150 |
| 5.4 | Relative Spectral Response for Different Cover Types | 151 |
| 5.5 | Weighted Average Spectral Response | 155 |
| 5.6 | Comparison of Edge and Weighted Acreage Responses | 155 |
| 5.7 | Results of LSD Test | 157 |
| 5.8 | Estimated Acreages ERTS and Percent Underestimation | 161 |
| 5.9 | <u>Water Resources Research</u> | |
| 6.1 | Percentage of Area Classification Using Eight Channels | 178 |
| 6.2 | Percentage of Area Classification Using Four Channels | 178 |
| | <u>Analysis Technique Development</u> | |
| 7.1 | Image Characteristics | 211 |
| 7.2 | Comparison of RIMPART and Per-Point Classifiers | 211 |
| 7.3 | Classification Results of Conventional and Binary Tree Procedures of Experiment I | 214 |
| 7.4 | Classification Results of Experiment II | 214 |
| | <u>Reformatting</u> | |
| 8.1 | Summary of ERTS CCT Frames Processed | 257 |
| 8.2 | Data Products Summary | 258 |
| 8.3 | Data Mean and Standard Deviation - Frame 1016-1605000 | 266 |
| 8.4 | CCT Data Mean and Standard Deviation Frame 1017-16093000 | 268 |
| 8.5 | Mean and RMS Across Track Misregistration Estimation | 269 |
| 8.6 | Mean and RMS Along Track Misregistration Estimation | 270 |

| | <u>Atmospheric Correction</u> | <u>Page</u> |
|------|---|-------------|
| 9.1 | Absorption Parameters in the Visible and Near-Infrared Spectrum | 301 |
| 9.2 | Absorption Coefficients and Rayleigh Scattering Optical Thickness for ERTS MSS Channels | 302 |
| 9.3 | Northern Illinois Reflectivities and Atmospheric Transmissivities | 309 |
| | <u>Comparison of System and Scene Corrected Data</u> | |
| 10.1 | Details of Data Used In Study | 312 |
| 10.2 | Means and Standard Deviations from Test Blocks in Scene Corrected and System Corrected MSS Data | 322 |

1.0 Introduction

This report describes research performed during the total contract period (July 1, 1972 - May 31, 1974) of the Purdue University-LARS ERTS-1 Wabash Valley Study. The study consists of nine projects as described in the Data Analysis Plan and progress and results are presented for each of these.

Section 2 presents progress and results from the Crop Species Identification Project. ERTS data from Indiana, Northern Illinois, Southeastern Missouri and Southeastern Idaho were analyzed to determine the accuracy with which major crop species could be identified using computer techniques. Section 3 discusses Soil Association Mapping project research which included computer classification of ERTS data and evaluation of results by overlaying soil association maps on the computer derived map. Section 4 presents results of the Urban Land Use Analysis Project. A detailed study of ERTS data from the Milwaukee, Wisconsin Marion County, Indiana and the Gary, Indiana area is described. Section 5 describes results for the Water Resources Research Project. ERTS-1 MSS data and ERIM aircraft scanner underflight data were analyzed to estimate the area of various water bodies. Look-sun angle effects are observed and discussed for the aircraft data. Section 6 describes Earth Surface Features Identification Project results. The test area for the study was re-classified using temporal data and improved agreement was achieved between ERTS results and the ground truth data for forest cover. Section 7 presents the results of the Analysis Technique Development Project. Section 8 describes Data Reformatting and Overlay. Section 9 presents results for the Atmospheric Modeling Project. Section 10 contains the results of

comparison of system corrected and scene corrected CCT data.

The work reported represents the final results of the research activities for the study.

2. IDENTIFICATION AND AREA ESTIMATION OF AGRICULTURAL CROPS BY COMPUTER CLASSIFICATION OF ERTS-1 MSS DATA

2.1 INTRODUCTION

Forecasting and estimating crop production is an activity of major importance practiced in most countries of the world. The value of this type of information is substantial. It is used in managing production, storage, transportation, and pricing of crops. Additionally, governments around the world use crop production information in designing national farm programs and establishing import-export policies. Remote sensing technology has the potential for significantly improving the quality of national and world crop production information.

Any improvement in the quality of this vital information could have far-reaching economic and social benefits. Eisgruber¹ shows that reduction from three to two percent in the error of estimate for corn, soybean, and wheat production estimates for the United States would result in 14 million dollars net social benefit to the country (1966-70 prices). Reduction to one percent would add another nine million dollars. On a world-wide basis, the value of improved crop production information would be magnified many times. Today, the value of improved information would be considerably greater because of the large increases in the prices of these commodities. More frequent and timely estimates alone, even without an accompanying improvement in accuracy, would result in additional benefits².

The wide area coverage of the ERTS, combined with the capabilities of computer data processing, offers a unique opportunity to reduce the above-mentioned error of estimate through reduction in sampling error. Furthermore, the

sequential coverage capabilities of ERTS may lead to benefits arising from improvements in the frequency and timeliness of estimates.

2.2 OBJECTIVES

Because of these potential benefits, the overall objective of this research has been to quantitatively evaluate the utility of the machine analysis of ERTS multispectral scanner (MSS) data in identifying crop species and estimating acreages. The investigations test the utility of ERTS MSS data in identifying crops over a range of environments with differing crops, soils, climates, and cultural practices. The effect on crop classification performance of a number of factors, including training sample size, extendability of training statistics, temporal and spatial dimensions of ERTS data, use of prior probability information in the classification algorithm, and wavelength band selection is examined.

The work consists of three major efforts: 1) classification of a three-county area in northern Illinois with 1972 ERTS data, 2) classification of 1973 ERTS data from two Indiana Crop Reporting Districts, and 3) a cooperative effort with the Statistical Reporting Service, U. S. Department of Agriculture in analyzing ERTS data from Missouri and Idaho.

2.3 ILLINOIS STUDY

The first ERTS data collected over the Corn Belt area available to LARS were frames 1017-16093 and 1017-16100 acquired August 9, 1972 (Figure 2.1). These frames were selected for analysis because no cloud-free ERTS coverage of Indiana was available for the 1972 growing season and because of its proximity to LARS for collecting ground truth data.

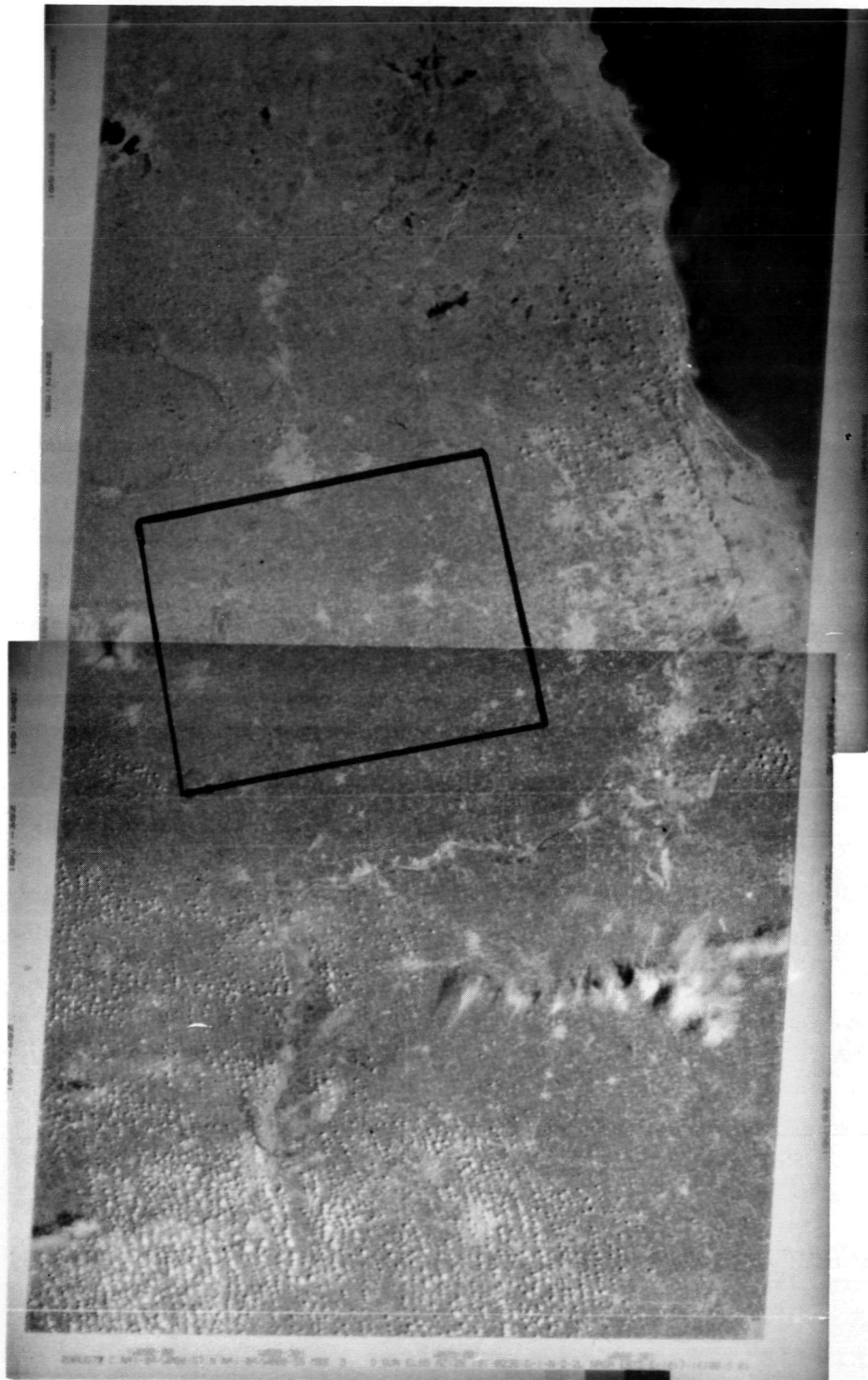


Figure 2.1 ERTS imagery (band 5, 0.6-0.7 μ m) of portions of frames 1017-093 and 1017-100 collected August 9, 1972 over northern Illinois. The three-county test site is outlined.

2.31 OBJECTIVES

The overall objective of this investigation, the first crop classification of ERTS data performed by LARS, was to quantitatively evaluate the potential of ERTS data for crop species identification and acreage estimation. Additional specific objectives were to:

1. Evaluate several analysis procedures and determine which was most effective for achieving accurate classification.
2. Determine the extendability and variability of training statistics.
3. Measure the effect of training set size on classification performance.
4. Evaluate the utility of several combinations of ERTS wavelength bands for classification.
5. Test the potential for improving classification performance by the inclusion of the temporal and/or spatial dimension of ERTS data in addition to the spectral dimension.
6. Develop and test methods of converting ERTS classifications to acreage estimates.

Classification performance was measured in two ways: (1) classification of test fields containing only "pure" pixels (commission/omission error matrix) and (2) comparison of area estimates derived from ERTS classifications to estimates made by the USDA.

2.32 PROCEDURES

A three-county area (DeKalb, Lee, and Ogle Counties) was selected for analysis. This area has highly productive, level to gently rolling soils and is intensively cropped. The primary crops grown are corn and soybeans; in 1972, about 60 percent of the total farmland was planted to these two crops (Tables 2.1 and 2.2).

Table 2.1 Estimated land use of DeKalb, Ogle, and Lee Counties, Illinois, 1967.*

| County | Total Land Area | <u>Non-Agricultural Land</u> | | <u>Agricultural Land</u> | | | |
|----------------|-----------------------|------------------------------|-------------------------|--------------------------|-------------|-------------|-------------|
| | | Urban and Built-Up | Small Water Areas | Cropland | Pasture | Forest | Other |
| Thousand Acres | | | | | | | |
| DeKalb | 407.0 | 15.0 | 0.5 | 372.8 | 4.1 | 5.7 | 8.9 |
| Ogle | 484.5 | 23.2 | 3.5 | 368.3 | 37.4 | 30.8 | 21.3 |
| Lee | <u>465.8</u> | <u>28.6</u> | <u>0.5</u> | <u>368.6</u> | <u>28.3</u> | <u>10.0</u> | <u>29.6</u> |
| TOTAL | 1,357.3 | 66.8 | 4.5 | 1,109.7 | 69.8 | 46.5 | 59.8 |

*Illinois Conservation Needs Inventory, 1967.

Note: This is the most quantitative and accurate description of land use available for the test area. And, it is also the most recent available information. While it is several years old the areas of the categories shown have not changed appreciably.

Table 2.2 USDA estimates of corn and soybean acreages in DeKalb, Ogle, and Lee Counties, Illinois, 1972.*

| County | Total Land | Corn | Soybeans | "Other" |
|----------------|--------------|--------------|--------------|--------------|
| Thousand Acres | | | | |
| DeKalb | 407.0 | 168.8 | 86.8 | 151.4 |
| Ogle | 484.5 | 200.1 | 55.8 | 228.6 |
| Lee | <u>465.8</u> | <u>176.6</u> | <u>102.1</u> | <u>187.1</u> |
| TOTAL | 1,357.3 | 545.5 | 244.7 | 567.1 |

*Illinois Agricultural Statistics, Annual Summary, 1973.

Note: On August 9 the corn and soybeans have both achieved their maximum vegetative growth. Neither crop has reached physiological maturity, however. Corn is fully tasseled by this date.

Ground truth data used to support the analysis consisted of identification by ground observation and recording on large-scale aerial photography of the crop or use of more than 500 fields in four different areas of the three counties. Crops identified were corn, soybeans, grain sorghum, and alfalfa; other cover types or land uses identified were hay, pasture, and small grain stubble, and woods. However, for the purposes of this analysis, only three classes were considered: corn, soybeans, and "other" (all cover other than corn and soybeans, including towns and highways). These 500 fields were used for training the maximum likelihood classifier and testing the accuracy of classifications.

Most fields could be accurately located in the ERTS data using a computer printout image generated on the basis of statistics from the nonsupervised classifier. A printout of band 5 was particularly useful for locating county and state highways, but individual fields could be found best in the clustered multiband imagery. Fields as small as about 10 acres could be located in the data, but pixels for training or testing the classifier were chosen only from fields larger than about 20 acres. After outlining the boundaries of fields on the imagery, coordinates of field centers containing only what was believed to be pure pixels (one crop or cover type only) were obtained.

Following the locating procedure, a random selection of training fields was made from each crop or cover type. All available fields not used for training were included in the test set. The number of corn and soybean training fields was varied from three to 12 in order to evaluate the effect of the number of training fields on classification performance. For the "other" classes, two to four fields of each cover type were included in the training set. Statistics for the selected set of training fields were then computed.

Classifications of each of the three counties were then carried out utilizing a training set comprised of fields from

that county. This led up to a classification of the entire three-county area which was made using training fields previously used for classifying the individual counties. Several additional analyses were also conducted to answer questions concerning the utility of the machine-processing of ERTS MSS data for crop species identification.

2.33 RESULTS AND DISCUSSION

A classification map of corn, soybeans, and "other" for a part of the three-county test area is shown in Figure 2.2. For crop inventory purposes, such maps are useful for qualitatively evaluating the classification. The large numbers of fields appearing as rectangles made-up of one class indicate that this is a good classification. More quantitative indicators of classification performance which were used to evaluate this classification and various factors influencing classification performance will be discussed in the remaining sections of this report.

2.331 CLASSIFICATION OF TEST FIELDS

The DeKalb County Area was classified first, using a training set with 12 corn fields, 12 soybean fields and two to three fields for each of five classes of "other". An overall test performance (total test points correctly classified/total test points classified) of 82.8% was achieved for the three classes of corn, soybeans, and "other". Quantitative results from this classification are shown in Table 2.3. Similar results obtained for the other two counties are presented in Tables 2.4 and 2.5.

The final classification of the entire three-county area was made by combining the training fields from each of the three counties, obtaining their class statistics, and classifying the entire test area. Two classifications, employing equal and unequal class weights, were performed. The use of prior probability information as class weights in the discriminate function is

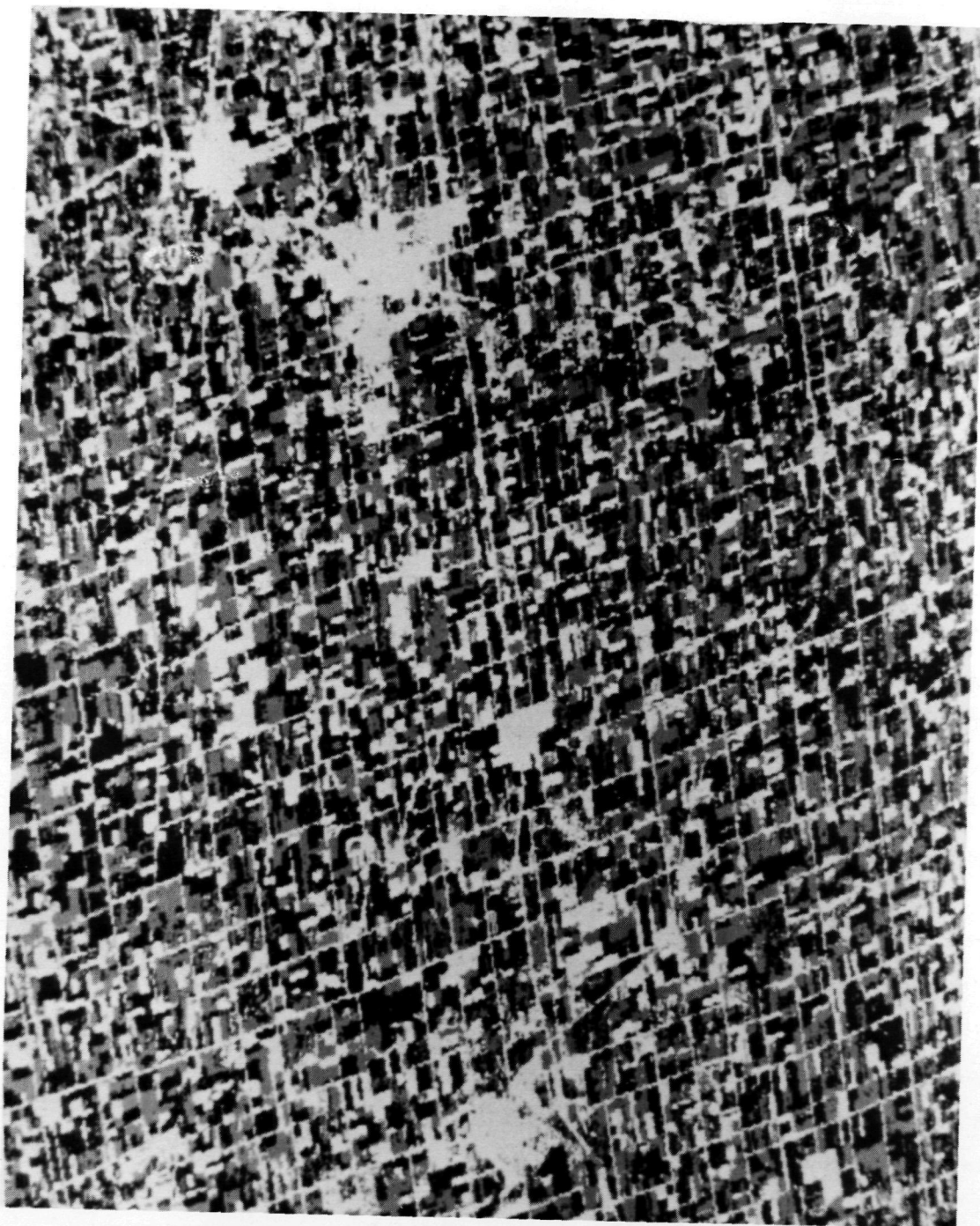


Figure 2.2. Computer classification CRT Image of corn, soybeans, and "other" for Dekalb, Ogle, and Lee Counties, Illinois. Corn-black, Soybeans-gray, and "other"-white.

Table 2.3 Classification of corn, soybean, and "other" test fields, DeKalb County, Illinois.

| CLASS | NO. POINTS | NO. POINTS CLASSIFIED AS | | | PERCENT CORRECTLY CLASSIFIED |
|----------|---------------|--------------------------|----------|---------|------------------------------------|
| | | CORN | SOYBEANS | "OTHER" | |
| Corn | 3968 | 3367 | 357 | 244 | 84.9 |
| Soybeans | 1113 | 115 | 855 | 133 | 76.8 |
| "Other" | 295 | 16 | 50 | 234 | 79.3 |
| TOTAL | 5376 | 3498 | 1262 | 611 | 82.8 |

Table 2.4 Classification of corn, soybean, and "other" test fields, Ogle County, Illinois.

| CLASS | NO. POINTS | NO. POINTS CLASSIFIED AS | | | PERCENT CORRECTLY CLASSIFIED |
|----------|---------------|--------------------------|----------|---------|------------------------------------|
| | | CORN | SOYBEANS | "OTHER" | |
| Corn | 3496 | 3016 | 59 | 424 | 86.3 |
| Soybeans | 629 | 158 | 394 | 77 | 62.6 |
| "Other" | 544 | 75 | 194 | 275 | 50.6 |
| TOTAL | 4669 | 3249 | 647 | 776 | 78.9 |

Table 2.5 Classification of corn, soybean, and "other" test fields, Lee County, Illinois.

| CLASS | NO. POINTS | NO. POINTS CLASSIFIED AS | | | PERCENT CORRECTLY CLASSIFIED |
|----------|---------------|--------------------------|----------|---------|------------------------------------|
| | | CORN | SOYBEANS | "OTHER" | |
| Corn | 1681 | 1460 | 60 | 161 | 86.9 |
| Soybeans | 552 | 131 | 389 | 32 | 70.5 |
| "Other" | 100 | 20 | 29 | 51 | 51.0 |
| TOTAL | 2333 | 1611 | 478 | 244 | 81.4 |

discussed in a later section of this chapter.

The distribution of test field classification performance is shown in Figure 2.3. The pixels in approximately 50% of the test fields were classified 91 to 100% correctly and 60% of the fields were identified at an accuracy of 80% or more. The next largest category was the 10 to 15% of the fields which were classified only zero to 10% correctly. Incorrect on-the-ground identifications of the crop in some of these fields is suspected since it was collected late in the season and the base photography on which fields were located was over a year old. Various factors affecting and accounting for this classification performance will be discussed in subsequent sections of this chapter.

Overall classification performance of approximately 80% is similar to or only slightly less than previously achieved with MSS data collected at 3000 meter altitudes or less. In the most extensive tests (15 flightlines, 24 km in length, six dates) by LARS of crop identification by computer classification of aircraft MSS data, the average correct recognition of corn vs. all other cover types was 85%³. These ERTS results are considered to be very positive considering the limited number of spectral bands, limited dynamic range, and gross spatial resolution of the ERTS data compared to aircraft scanner data. ERTS data, however, does not have the problems with sun and view angles which have been present in aircraft scanner data. The results are particularly significant because they were from a 3000 square km area compared to classifications of aircraft data which generally were only about 20 square km. Qualitative checks of classification maps of areas outside the three-county area indicated that classification performance was probably satisfactory over a considerably larger area.

2.332 USE OF A PRIORI PROBABILITIES IN CLASSIFICATION

A primary objective of this investigation has been to determine the applicability of available data analysis procedures

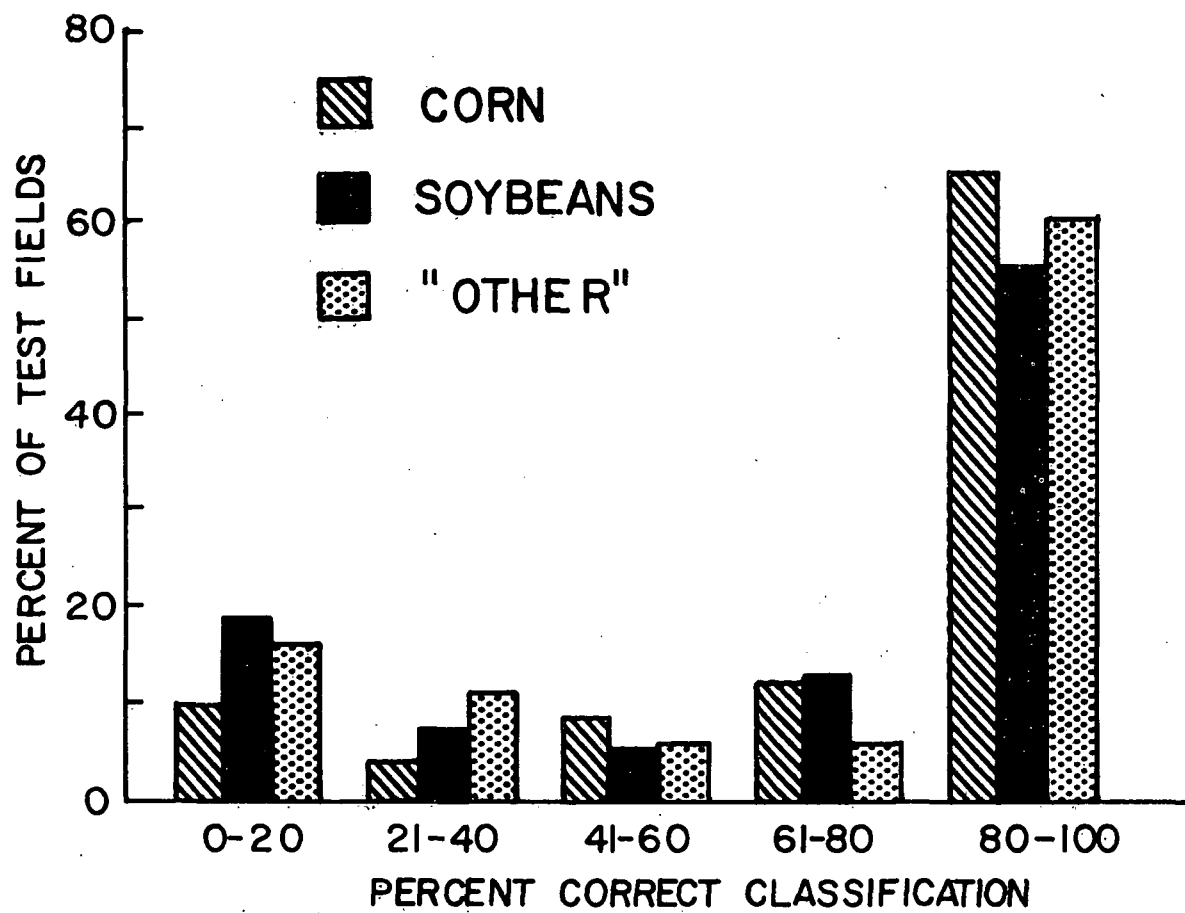


Figure 2.3. Distribution of test field classification performance.

to crop identification and acreage estimation with ERTS MSS data. The use of *a priori* probability information in the classification decision rule and unbiasing of classification results are two procedures which we believe should be particularly useful for improving crop acreage estimates based on classifications of ERTS data. Results from these two procedures will be discussed in this and the next section.

This ERTS investigation provided the impetus for implementing a capability to utilize prior probability information or class weights in the classifier's decision rule. Prior to this time equal probabilities of occurrence had been assumed for all classes. It was known, however, that in certain cases the assumed equal class weights were far from the true situation. The inclusion of prior probability information in the decision rule makes it Bayes optimal; that is, minimizes the probability of classifying an object into class i when it is actually from class j .

In the case of agricultural crops, possible sources of prior probability information are: estimates from the previous year, an earlier survey or classification this year, or a survey of farmers' planting intentions. In our current work, we have used USDA/SRS acreage estimates from the previous year (1971) as weights. The classification results for the test fields from the three counties with unequal class weights are presented in Table 2.6b. All previous results shown have been for the equal class weight case. These results should be compared to those in Table 2.6a. The weights were 44, 16, and 40 for corn, soybeans, and "other", respectively; these are the estimated percentages of each class present in the three-county area in 1971.

The accuracy of corn recognition was increased about 5% and "other" increased 2.5%, but soybean recognition decreased almost 8%. Overall classification performance was increased 2.3%. Our conclusion from these limited results is that while the theoretical

Table 2.6 Classification of corn, soybean, and "other" test fields, DeKalb, Ogle, and Lee Counties, Illinois, with and without the use of prior probability information in the classification decision rule.

(a) No prior probability information used, equal class weights assumed.

| CLASS | NO. POINTS | NO. POINTS CLASSIFIED AS | | | PERCENT CORRECTLY CLASSIFIED |
|----------|---------------|--------------------------|----------|---------|------------------------------------|
| | | CORN | SOYBEANS | "OTHER" | |
| Corn | 9290 | 7546 | 973 | 771 | 81.2 |
| Soybeans | 2235 | 244 | 1732 | 259 | 77.5 |
| "Other" | 1121 | 150 | 307 | 664 | 59.2 |
| TOTAL | 12646 | 7940 | 3012 | 1694 | 78.6 |

(b) Prior probability information used, unequal class weights.*

| CLASS | NO. POINTS | NO. POINTS CLASSIFIED AS | | | PERCENT CORRECTLY CLASSIFIED |
|----------|---------------|--------------------------|----------|---------|------------------------------------|
| | | CORN | SOYBEANS | "OTHER" | |
| Corn | 9290 | 7983 | 382 | 925 | 85.9 |
| Soybeans | 2235 | 395 | 1556 | 284 | 69.6 |
| "Other" | 1121 | 206 | 220 | 695 | 62.0 |
| TOTAL | 12646 | 8584 | 2158 | 1904 | 80.9 |

*Class weights were 44, 16, and 40 for corn, soybeans, and "other", respectively.

basis for the use of prior probability information in the classification discriminate function is sound, in practice it may not significantly change the results as appears to be the case here. Greater improvement could be expected when the class weights differ more from the equal weights than these did. More experiments to test the use of prior probability information in classification need to be performed.

2.333 UNBIASING CLASSIFICATION RESULTS

Experience has shown that it is inevitable that some points are incorrectly identified by the maximum likelihood classifier. In this experiment, only about 80% of the test samples were correctly classified. The primary source of these errors is overlapping density functions for two or more classes. For example, some corn "looks" like soybeans and some soybeans are spectrally similar to corn. As described above, prior probability information or class weights can be used to good advantage to at least partially reduce the effects of such circumstances. A second procedure which can be used after the classification has been performed is to unbias or adjust the results based on the correct classification proportions and error rates. The latter procedure was first used during the 1971 Corn Blight Watch Experiment³.

The source of the correct classification proportions and error rates are the matrices of test field classification performance such as shown in Table 2.6. From such information, we can determine the proportions, for instance, of corn classified as corn and non-corn and the proportions of non-corn classified as non-corn and corn. With this information, it is then possible to unbias or adjust the classification results for a county or several counties so that they more nearly estimate the true amounts of each class present in the classified area.

Theoretically, if the true values of the error rates of omission and commission were known, the classification results could

be adjusted so that in effect the area estimates based on the classification closely approximated the true amounts of each crop present. In practice, of course, this situation is seldom found. The primary limitation is that the test samples are not completely representative of the total area classified and only provide estimates of the true error rates. Possible causes of non-representative test samples are that samples come from only a small part of the total area being classified and that many cover types such as farmsteads, idle land, roads, and urban areas generally have not been included in the set of test fields.

The method we have used for unbiasing classification results involves multiplying the county classification results (Table 2.7) by the inverse of the test field classification performance matrix (Table 2.6) as follows:

$$A = CP^{-1}$$

where, C is the classification vector with n crops or classes, P^{-1} is the inverse of the n x n classification performance matrix, and A is a 1 x n vector of the crop acreages.

The results of applying this correction procedure are presented in Table 2.8 and discussed in the next section along with further results on the use of prior probabilities to the classification function.

2.334 ACREAGE ESTIMATION

The classification performance indicated by 80% correct recognition of test fields is believed to be adequate for satisfactorily estimating crop acreages. To determine how well crop acreages could be estimated from the ERTS classification, the ERTS coordinates of the three counties were obtained, the counties were classified, and the number of pixels classified

Table 2.7 Number of samples classified into corn, soybeans, and "other" for DeKalb, Ogle, and Lee Counties, Illinois.

(a) Equal Class Weights

| County | No. Points Classified As | | |
|--------|--------------------------|----------|---------|
| | Corn | Soybeans | "Other" |
| DeKalb | 131,451 | 85,148 | 74,311 |
| Ogle | 146,108 | 112,385 | 135,058 |
| Lee | 150,992 | 122,101 | 120,266 |
| TOTAL | 428,551 | 319,634 | 329,635 |

(b) Unequal Class Weights

| County | No. Points Classified As | | |
|--------|--------------------------|----------|---------|
| | Corn | Soybeans | "Other" |
| DeKalb | 152,920 | 54,948 | 83,042 |
| Ogle | 170,220 | 74,940 | 148,391 |
| Lee | 178,177 | 80,241 | 134,941 |
| TOTAL | 501,317 | 210,129 | 366,374 |

into each class tabulated (Table 2.7). In Table 2.8, four acreage estimates based on the ERTS classifications are compared to each other and to estimates made by the Illinois Cooperative Crop Reporting Service (SRS/USDA). The ERTS estimates are the four combinations of using prior probability information in the classification decision rule and unbiasing the classification results as discussed in the previous two sections.

The standard to which the ERTS classifications are compared is the acreage estimates (shown as percentage of total land area) made by SRS/USDA. The mean squared differences between the SRS/USDA estimates and the several ERTS estimates are shown as a means of comparing the overall goodness of each ERTS estimate.

One of the most difficult aspects of remote sensing technology is quantitatively evaluating classification results. It is physically impossible to collect sufficient ground data of crop identification and acreage over large areas, to determine how accurate area estimates made from the ERTS classification are. We have therefore used the USDA county estimates as the reference for comparison. However, the crop surveys conducted by the USDA are designed to achieve prescribed levels of accuracy at only the national and state levels. For this reason the USDA does not publish accuracy figures for their county estimates. However, in those states, including Illinois, in which an annual farm census is conducted, the acreage estimates are considered to be quite accurate. Their estimates are probably within three to five percent of the actual acreages.

The ERTS estimates, particularly those adjusted for classification bias, are very close to those made by the USDA. It seems clear that the USDA and ERTS estimates are of the same parameter. The estimates agree best for the total of the three counties. There is more variation between the two estimates for the individual counties. However, this is simply a result of having a larger sample and can be expected as long as there is not a consistent bias in one

Table 2.8. Comparison of crop acreage estimates by USDA and estimates based on ERTS classifications. The results of utilizing prior probability information in classification and bias correction of classifications are shown.

| County | Class | SRS- USDA | ERTS | | | |
|------------------------------|----------|--------------|---------------|-----------------|----------------|-----------------|
| | | | Uncorrected | | Bias Corrected | |
| | | | Equal Wts. | Non-Eq. Wts. | Equal Wts. | Non-Eq. Wts. |
| (Percent of Total Land Area) | | | | | | |
| DeKalb | Corn | 41.5 | 45.2 | 52.6 | 47.6 | 50.8 |
| | Soybeans | 21.3 | 29.3 | 18.9 | 19.9 | 14.3 |
| | "Other" | 37.2 | 25.5 | 28.5 | 32.5 | 34.9 |
| | r.m.s.* | | 6.5 | 6.9 | 8.3 | 8.5 |
| Ogle | Corn | 41.3 | 37.1 | 43.3 | 35.5 | 37.0 |
| | Soybeans | 11.5 | 28.6 | 19.0 | 14.4 | 10.2 |
| | "Other" | 47.2 | 34.3 | 37.7 | 50.1 | 52.8 |
| | r.m.s. | | 12.6 | 7.1 | 4.1 | 4.1 |
| Lee | Corn | 37.9 | 38.4 | 45.3 | 37.6 | 40.0 |
| | Soybeans | 21.9 | 31.0 | 20.4 | 19.9 | 14.0 |
| | "Other" | 40.2 | 30.6 | 34.3 | 42.5 | 46.0 |
| | r.m.s. | | 7.6 | 5.5 | 1.8 | 5.8 |
| Total for all Counties | Corn | 40.2 | 39.8 | 46.5 | 39.6 | 41.8 |
| | Soybeans | 18.0 | 29.6 | 19.5 | 17.8 | 12.7 |
| | "Other" | 41.8 | 30.6 | 34.0 | 42.6 | 45.5 |
| | r.m.s. | | 9.3 | 5.8 | 0.6 | 3.9 |

*r.m.s.-root mean square difference between USDA and ERTS estimates.

direction in the ERTS classification, e.g. corn is always overestimated.

Two additional things are clear from the results: (1) the use of class weights or prior probability information in classification gave substantially better estimates of the amounts of corn and soybeans present (reduction of the r.m.s. difference from 9.3 to 5.8) and (2) the application of the unbiasing procedure after classification further improved the ERTS estimates (reduction of r.m.s. difference to 3.9 and 0.6) for the classification with and without class weights, respectively. In conclusion, these two procedures should be used whenever possible in making crop acreage estimates from classifications of ERTS type data.

2.335 EXTENDABILITY OF TRAINING STATISTICS

The extendability, variability, and size of training sets required to achieve accurate classifications have a large impact on the design and requirements of a ground observation system to use with remote sensing. To test extendability of training sets, the training set from each county was also used to classify the other two counties. Results (Figure 2.4) show that equally good performance was achieved by using any of the three training sets for classifying the three areas, which are 24 to 40 km (15 to 25 miles) apart. Qualitative examination of classification maps such as in Figure 2.2 indicate that "good" classifications were produced from these same statistics up to 80 km from the origin of the statistics.

The distance which statistics can be successfully extended will depend on a number of factors. An important requirement will be that the composition of the ground scene is similar to that where the training statistics were obtained. If the cover types or their condition changes significantly, then the statistics will probably not be valid. Atmospheric conditions should also be nearly constant. The ability to extend statistics will be improved as the capability to radiometrically calibrate satellite

Training Set Extendability

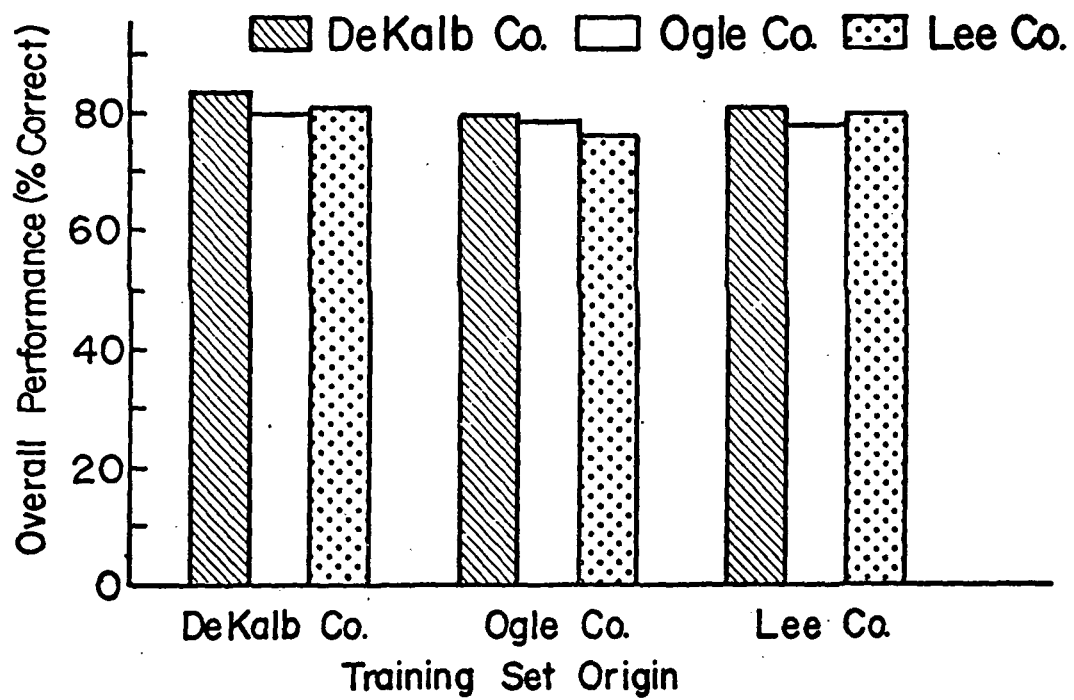


Figure 2.4. Test of the extendability of training sets.

data is improved. When these conditions are met, it may be possible to extend signatures large distances, i.e., to other ERTS frames; however, considerably more research will be required to develop the capability to extend statistics beyond a few frames.

2.336 NUMBER OF TRAINING FIELDS

An important question in developing remote sensing technology is how much ground observation data is required for successfully classifying an area. To be cost-effective, the amount of ground data should be only a small fraction of that required for conducting a conventional ground-level survey. To begin to accumulate evidence to answer this question, classifications were performed with varying amounts of ground observation data used for training the classifier.

Initially, 12 corn and 12 soybean fields, randomly selected, were used to classify each test site. Subsequently, classification with nine, six, and three corn and soybean training fields were performed. Results of this test are shown in Figure 2.5. Although there was a small decrease in the accuracy of corn identification in the reduction from 12 to nine fields, the remainder of the training sets for corn and the three-, six-, and nine-soybean training field sets performed as well as the 12.

It should be pointed out that while these results indicate that only a small number of training fields was required in this particular instance, more training data might be needed in those areas having more variability in crops. On the other hand, as we learn more about the spectral characteristics of crops and other cover types, particularly in relation to their growth and development (crop calendar), it may be possible to reliably identify major crop species directly from ERTS imagery for training purposes making routine ground observations unnecessary. This would be a particularly valuable capability for inaccessible areas.

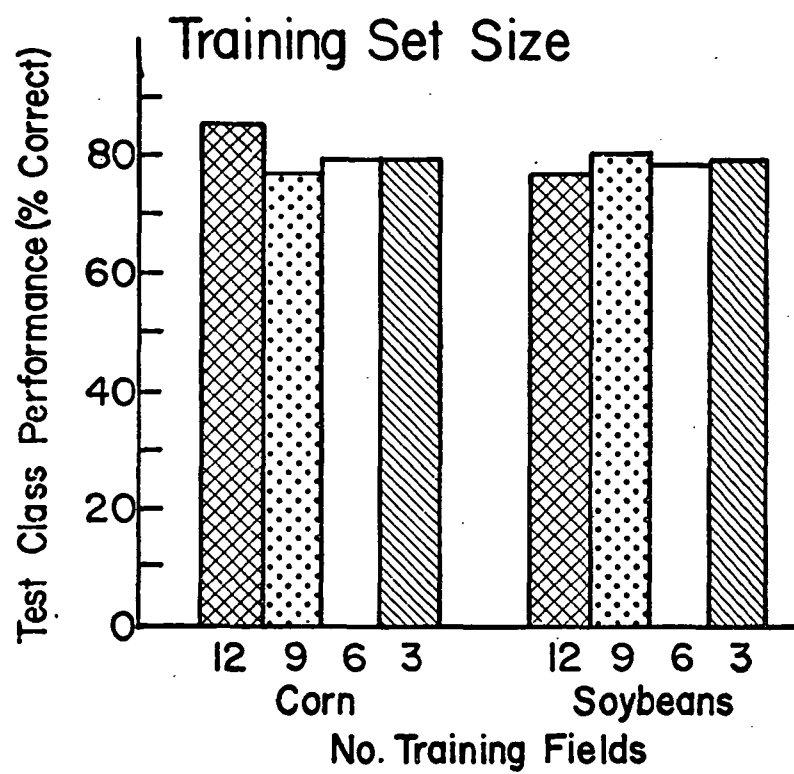


Figure 2.5. Influence of training set size on classification performance.

2.337 VARIABILITY OF CLASSIFICATION RESULTS

The number of training fields required to satisfactorily classify an area is directly related to the variation in spectral response exhibited by the cover types of interest. The more variation there is within a particular cover type, the more fields will be required to adequately represent it. The results on the number and extendability of training fields indicate that little variation in classification performance would result from using different sets of training fields. To test this hypothesis, four random selections of training sets were made from each county. Each training set consisted of 12 fields each of corn, soybeans, and other.

Test field classification performances are summarized in Table 2.9. Recognition of corn was very stable both within and among the three test counties. The coefficient of variation (CV) for corn was 1.6%. Soybeans displayed a greater amount of variation (CV=17.4). However, most of the variation was due to one of Ogle County training sets which resulted in very poor recognition of soybeans. The CVs for soybeans in DeKalb and Ogle Counties were both 6.8%. The recognition of "other" was considerably more variable than for either corn or soybeans. This is as expected because of the inclusion of several different cover types within the class of "other". In fact, hay or pasture crops can be extremely variable because these categories may contain fields of several different species and be in any of several stages of maturity.

The results show that a reasonably large random sample of corn and soybean training fields adequately represented these classes, but that more training fields would be required to satisfactorily represent the class of "other". These results are supported by the variation in spectral response of the various cover types to be discussed in the next section.

Table 2.9. Variation in test field classification performance due to training field selection.

| County | Training Set | Test Field Classification (% Correct) | | | |
|--------------------|--------------|---------------------------------------|---------|---------|---------|
| | | Corn | Soybean | "Other" | Overall |
| DeKalb | 1 | 84.9 | 76.8 | 79.3 | 82.8 |
| | 2 | 85.4 | 68.4 | 77.9 | 81.4 |
| | 3 | 87.7 | 69.5 | 74.5 | 83.5 |
| | 4 | 84.8 | 78.1 | 80.2 | 83.2 |
| | Mean | 85.7 | 73.2 | 77.97 | 82.7 |
| | Std. Dev. | 1.4 | 5.0 | 2.5 | 0.9 |
| | C. V. (%) | 1.6 | 6.8 | 3.2 | 1.1 |
| Ogle | 1 | 83.3 | 33.6 | 81.3 | 77.6 |
| | 2 | 86.3 | 62.6 | 50.6 | 78.6 |
| | 3 | 86.5 | 72.9 | 30.2 | 78.1 |
| | 4 | 87.3 | 74.1 | 24.8 | 77.9 |
| | Mean | 85.8 | 60.8 | 46.7 | 78.0 |
| | Std. Dev. | 1.8 | 18.9 | 25.6 | 0.4 |
| | C. V. (%) | 2.1 | 31.1 | 54.8 | 0.5 |
| Lee | 1 | 86.9 | 70.5 | 51.0 | 81.4 |
| | 2 | 85.4 | 63.3 | 26.1 | 75.1 |
| | 3 | 84.7 | 64.1 | 46.3 | 77.9 |
| | 4 | 88.1 | 72.6 | 29.0 | 80.1 |
| | Mean | 86.2 | 67.62 | 38.1 | 78.6 |
| | Std. Dev. | 1.5 | 4.6 | 12.4 | 2.8 |
| | C. V. (%) | 1.7 | 6.8 | 32.5 | 3.6 |
| Grand Mean Overall | | 85.9 | 67.21 | 54.2 | 79.8 |
| Std. Dev. | | 1.4 | 11.7 | 23.3 | 2.7 |
| C.V. (%) | | 1.6 | 17.4 | 42.9 | 3.4 |

C.V. - Coefficient of Variation

2.338 WAVELENGTH BAND SELECTION

Most ERTS classifications have been performed using all four ERTS MSS bands. This is probably a carryover from analysis of aircraft scanner data which showed that maximum classification accuracy was approached using four to six of the 12 or more available channels. In an operational system, it will be important to use the set of channels which provides the optimal trade-off between classification costs (complexity and computation time) and classification accuracy. The objectives of this study were: (1) determine the informational value of the various ERTS bands for crop identification, and (2) determine whether all four ERTS bands were required to maximize classification performance.

Since it was impractical to perform classifications of all combinations of the four ERTS MSS channels, the separability of the classes was estimated from their transformed divergence.* Divergence is a measure of the dissimilarity of two distributions and provides an indirect measure of the ability to discriminate between them. The results are presented in Table 2.10 along with classification accuracies for selected combinations of ERTS bands.

The first conclusion drawn from these results is that one ERTS band alone would be inadequate for satisfactorily identifying crop species. However, the divergence values as well as classification performances strongly suggest that the combination of one visible band and one near-infrared band results in crop classification accuracies as high as those obtained when three or four bands were included. This result does not, however, mean that ERTS has too many spectral bands. Rather, for this particular situation two bands contained essentially all the information required to discriminate among crops present. Working with aircraft scanner data having 12

Table 2.10 Interclass divergence for all combinations of ERTS bands and classification performance for selected combinations of bands.

| ERTS Channels* | Interclass Divergence | | | | Average Divergence | Classification Accuracy (%) | | | |
|-------------------|-----------------------|------|------|--|-----------------------|-----------------------------|---------|---------|---------|
| | CS | CO | SO | | | Corn | Soybean | "Other" | Overall |
| 5 | 109 | 1976 | 1674 | | 1253 | | | | |
| 4 | 187 | 1510 | 1195 | | 964 | | | | |
| 6 | 1193 | 510 | 314 | | 672 | | | | |
| 7 | 949 | 566 | 367 | | 627 | | | | |
| 5,6 | 1244 | 1982 | 1705 | | 1644 | 85.2 | 76.0 | 82.0 | 82.6 |
| 5,7 | 1029 | 1981 | 1703 | | 1571 | 83.9 | 71.7 | 76.6 | 81.0 |
| 4,6 | 1212 | 1653 | 1448 | | 1438 | | | | |
| 4,7 | 1002 | 1675 | 1450 | | 1376 | | | | |
| 4,5 | 268 | 1979 | 1711 | | 1319 | | | | |
| 6,7 | 1306 | 1059 | 411 | | 925 | | | | |
| 5,6,7 | 1338 | 1984 | 1709 | | 1677 | 83.2 | 78.0 | 78.6 | 81.3 |
| 4,5,6 | 1250 | 1984 | 1744 | | 1659 | | | | |
| 4,5,7 | 1051 | 1983 | 1742 | | 1592 | | | | |
| 4,6,7 | 1313 | 1732 | 1458 | | 1501 | | | | |
| 4,5,6,7 | 1343 | 1986 | 1747 | | 1692 | 82.1 | 79.9 | 81.6 | 81.1 |

* Ranked in order of average divergence

or more channels in the 0.40 to 15.0 μm range, Kumar and Silva found that when four channels were to be selected, one channel each from the visible near-, middle-, and thermal-infrared gave the highest separability.

While these limited results may not apply to all locations and conditions, they indicate the ERTS-1 wavelength bands may not be optimum for crop identification since all of the information required to obtain the highest classification performance was contained in two of the four bands. Improved classification performance could probably be expected if bands were available from the middle and thermal infrared.

2.339 SPECTRAL CHARACTERISTICS

When ground truth for training the classifier is available, it is possible to successfully classify MSS data with little consideration of the spectral characteristics of the cover types involved. However, it is important to know as fully as possible their reflectance properties in order to understand why the classifier was or was not able to discriminate among various materials. And, more importantly, if the reflectance characteristics of crops (and neighboring cover types, as well) in relation to their growth and development are known then it may be possible that crop identifications could be made without the use of ground observation for training. Such a capability would increase considerably the utility and value of the technology.

With these considerations in mind, the mean and variance of several cover types was determined. The approach was to cluster each of the major classes (corn, soybeans, grain sorghum, pasture, hay, and miscellaneous) to isolate spectral subclasses with unimodal, approximately Gaussian distributions. The criteria for determining that the clusters were distinct is described by Davis and Swain.⁶ The results are presented in Table 2.11.

Table 2.11 Mean and standard deviation of the relative spectral response of corn, soybeans, and other cover types.

| Class | Mean | | | | Standard Deviation | | | |
|-----------|-----------------------------------|---------------|---------------|---------------|-----------------------------------|---------------|---------------|---------------|
| | Wavelength Band (μm) | | | | Wavelength Band (μm) | | | |
| | 0.50- 0.60 | 0.60- 0.70 | 0.70- 0.80 | 0.80- 1.10 | 0.50- 0.60 | 0.60- 0.70 | 0.70- 0.80 | 0.80- 1.10 |
| Corn | 22.0 | 13.2 | 46.9 | 30.4 | 1.18 | 1.11 | 4.92 | 3.47 |
| Soybeans | 23.1 | 13.6 | 61.9 | 38.5 | 1.26 | 1.66 | 11.53 | 8.18 |
| Hay 1 | 26.2 | 15.2 | 79.4 | 48.1 | 2.52 | 2.20 | 6.49 | 3.07 |
| Hay 2 | 24.1 | 15.4 | 58.9 | 35.8 | 1.68 | 2.66 | 5.31 | 3.58 |
| Hay 3 | 33.3 | 34.0 | 47.0 | 23.9 | 2.27 | 4.23 | 4.99 | 3.08 |
| Hay 4 | 26.9 | 23.1 | 42.2 | 22.9 | 2.17 | 3.48 | 5.50 | 3.87 |
| Hay 5 | 22.2 | 16.5 | 21.5 | 10.9 | 1.81 | 2.87 | 4.75 | 2.99 |
| Pasture 1 | 25.8 | 18.0 | 53.7 | 32.1 | 1.67 | 2.29 | 2.46 | 1.81 |
| Pasture 2 | 28.1 | 23.3 | 47.4 | 26.9 | 2.30 | 3.86 | 3.44 | 1.94 |
| Pasture 3 | 22.7 | 17.5 | 20.9 | 10.3 | 1.10 | 1.55 | 3.18 | 2.23 |
| Sorghum | 25.0 | 19.2 | 39.7 | 22.4 | 1.22 | 2.27 | 7.57 | 5.76 |
| Woods | 22.9 | 14.8 | 47.9 | 29.5 | 2.04 | 2.56 | 3.54 | 2.35 |
| Misc. 1 | 27.1 | 20.1 | 58.3 | 34.5 | 4.29 | 7.79 | 8.67 | 7.26 |
| Misc. 2 | 25.6 | 18.9 | 46.6 | 27.1 | 4.81 | 7.37 | 4.99 | 4.77 |
| Misc. 3 | 23.8 | 19.8 | 24.9 | 12.5 | 2.13 | 3.13 | 5.66 | 3.27 |

As mentioned earlier, the major crop classes, corn and soybeans, as well as grain sorghum and woods were relatively uniform. There was only one spectral class for each of these cover types compared to three or more spectral subclasses for hay, pasture, and miscellaneous. This explains why classification performance was maintained at a high level even when the number of corn and soybean training fields was reduced to as few as three. It also indicates that a much larger number of fields of the other cover types would be required to adequately represent them.

The mean relative response of corn and soybeans as well as several of the other classes was similar in the two visible bands, but different in the infrared bands. The reflectance of soybeans was greater than for corn in the infrared bands. However, the response of corn and woods was nearly the same in all bands. Soybeans and hay, primarily alfalfa, exhibited similar responses in all bands. These were also the major sources of confusion in the classifications.

2.340 UTILIZATION OF THE SPATIAL DIMENSION OF ERTS DATA

Machine analysis of remote sensor imagery from aircraft and satellite sensors has primarily utilized the spectral measurement dimension. In spectral analysis, the scene is examined in terms of its reflectivity or emissivity. The spectral measurements are essentially instantaneous in time for each scene element. Two other forms of measurements can be made in conjunction with the measurement of reflected energy. The shape or spatial structure of scene objects can be observed and utilized to aid in extracting information from the measurements. Such analysis implies an imaging sensor which measures energy in a two-dimensional format with received energy being spatially resolvable to some given level. The second form is measurement of temporal variations and refers

to the observation of reflected or radiated energy as a function of time. The use of these two dimensions of ERTS data will be discussed in this and the following section.

Utilization of spatial structure in remote sensor imagery requires effective feature extraction procedures to derive size, shape, and textural characteristics from observed scenes. The sample classify function of LARSYS provides an experimental system for testing a classification scheme which uses spatial information as well as spectral information. It employs mean vectors and covariance matrices for the training classes and test fields to calculate the probability density function of each test field being classified and the probability density function of each of the training classes. It then assigns the field (rather than an individual data point) to the closest training class. The DeKalb County test fields were classified in this manner.

Recognition accuracy (Table 2.12) for corn and soybeans increased about five percent compared to the point classification previously shown in Table 2.3. This preliminary result indicates some improvement in classification performance might be obtained by utilizing the spatial dimension of ERTS data. Currently, however, it is impractical to use this classifier because field boundary coordinates are required as input to the classifier. In an operational survey, of course, location of field boundaries in the ERTS data would not be known. There has been, however, some research on the development of algorithms which determine where field boundaries occur in the data.⁷ This approach has been more successful with aircraft scanner data than with ERTS data because of the small number of points in many fields in the ERTS data. It does offer promise in those regions having many large fields or when satellite data collection systems with greater spatial resolution become available.

2.341 UTILIZATION OF THE TEMPORAL DIMENSION OF ERTS DATA

The use of temporal data in the context of remote sensing is relatively new and there are a few published reports of its

Table 2.12 Comparison of classification performance (percent correct recognition) for sample (field) and point classifications of DeKalb County test fields.

| Class | No. Fields | No. Points | Sample Classification | | Point Classification |
|----------|---------------|---------------|--------------------------|-------------|-------------------------|
| | | | Fields | Points | |
| Corn | <u>124</u> | 3,968 | 89.5 | 88.5 | 84.9 |
| Soybeans | <u>53</u> | 1,113 | 81.1 | 79.4 | 76.8 |
| "Other" | <u>24</u> | <u>295</u> | <u>70.8</u> | <u>70.8</u> | <u>79.3</u> |
| TOTAL | 201 | 5,376 | 84.1 | 85.6 | 82.8 |

use for crop identification.⁸ Nevertheless, it is generally believed that the temporal dimension has considerable potential for improving the recognition of crop species. It should be particularly beneficial in those cases where at any single time two or more crops may be inseparable spectrally, but with the addition of information on temporal changes in their spectral response may become discriminable. Observation of temporal variations requires, then, that spectral measurements be repeated continuously or at discrete intervals. ERTS has such a capability.

Cloud-free ERTS coverage of portions of DeKalb and Ogle counties was acquired on August 9, September 19, and October 2, 1972, and were registered to achieve geometric coincidence of the image elements of the three dates (see reference 9 for a description of the registration process). The new channels added by the repeated coverage of the same area were treated as additional spectral channels and used in the same way as the spectral channels. Pattern recognition techniques can then be applied to the expanded measurement space and the possible benefits of the added channels evaluated using existing multi-spectral analysis techniques.

The three time ERTS spectral-temporal data set offers 12 dimensions for analysis. All or any subset of these dimensions, or channels, can be used to recognize the cover types in the scene. Classification results using only spectral data are compared with results using both spectral and temporal data as the method of evaluating the temporal dimension. Results for classifications utilizing only the spectral dimension are shown in Table 2.13.

The best results were obtained for the August 9 data. After the end of August maturation begins to occur and all crops start moving spectrally toward a uniform brown color. In addition, the lower sun angle will probably cause any spectral differences among cover types to be smaller. As a result, their spectral separability decreases. The average interclass

Table 2.13 Comparison of training and test field classification performance for August 9, September 19, and October 2.

| CLASS | TRAINING FIELDS | | | TEST FIELDS | | |
|-----------|-----------------|---------|--------|-------------|----------|--------|
| | AUG. 9 | SEPT.19 | OCT. 2 | AUG. 9 | SEPT. 19 | OCT. 2 |
| % CORRECT | | | | | | |
| Corn | 87.3 | 80.0 | 79.8 | 76.3 | 56.6 | 53.8 |
| Soybeans | 90.6 | 63.8 | 71.0 | 84.9 | 52.5 | 48.3 |
| "Other" | 94.0 | 53.9 | 87.6 | 64.6 | 54.7 | 71.6 |
| Overall | 89.9 | 69.1 | 79.5 | 75.9 | 55.9 | 55.3 |

Table 2.14 Mean spectral response of corn and soybeans as a function of wavelength band and date.

| ERTS BAND | AUGUST 9 | | SEPTEMBER 14 | | OCTOBER 2 | |
|--------------|----------|----------|--------------|----------|-----------|----------|
| | CORN | SOYBEANS | CORN | SOYBEANS | CORN | SOYBEANS |
| 4 | 22.3 | 23.0 | 20.3 | 21.8 | 23.6 | 24.5 |
| 5 | 13.8 | 13.4 | 15.5 | 16.2 | 19.5 | 21.8 |
| 6 | 47.3 | 65.4 | 35.3 | 50.3 | 26.8 | 26.9 |
| 7 | 30.5 | 41.2 | 23.1 | 31.8 | 14.1 | 14.3 |

divergence dropped from 1879 on August 9 to 1147 and 1220 on September 19 and October 2, respectively. Classification accuracy also decreased between August and the two later dates.

The temporal changes in spectral response can be shown graphically by plotting the mean relative spectral response for a large number of samples for the three dates. The mean spectral response of corn and soybeans for the three dates are plotted in Figure 2.6. The steadily decreasing infrared reflectance (.7-.8 μm and .8-1.1 μm bands) shows the loss of green leaf vigor due to drying, decreased leaf area and ground cover; the increasing red wavelength band (.6-.7 μm) values indicate the loss of chlorophyll absorption as the plants mature and dry. Temporal effects on separability can also be illustrated by plotting mean spectral response for two bands with time as a parameter. Figure 2.7 contains such a response "trajectory" for the corn and soybeans training samples which shows dramatically how the distance between the classes decreases with time. Table 2.14 contains the mean spectral responses for the two classes. The Euclidian distance computed for bands 5 and 6 between the two classes is 18.2, 15.2, and 2.3 for August 9, September 19, and October 2, respectively. In summary, these several ways of looking at the temporal-spectral data show that the August data is superior to that from September or October for separating corn and soybeans.

Next, the twelve spectral-temporal channels were treated as one set of measurements and used to classify the same training and test samples used in the spectral experiments. Choosing the channels to use becomes the critical question since twelve channel classifications are excessively costly for large areas and there are a total of 4095 subsets of the twelve channels (i.e. all combinations of 1, 2, ... 12 channels) to choose from. The LARSYS feature selection program was employed to select the best subset of four channels from the twelve available for spectral-temporal classification. First, the best four out of the twelve

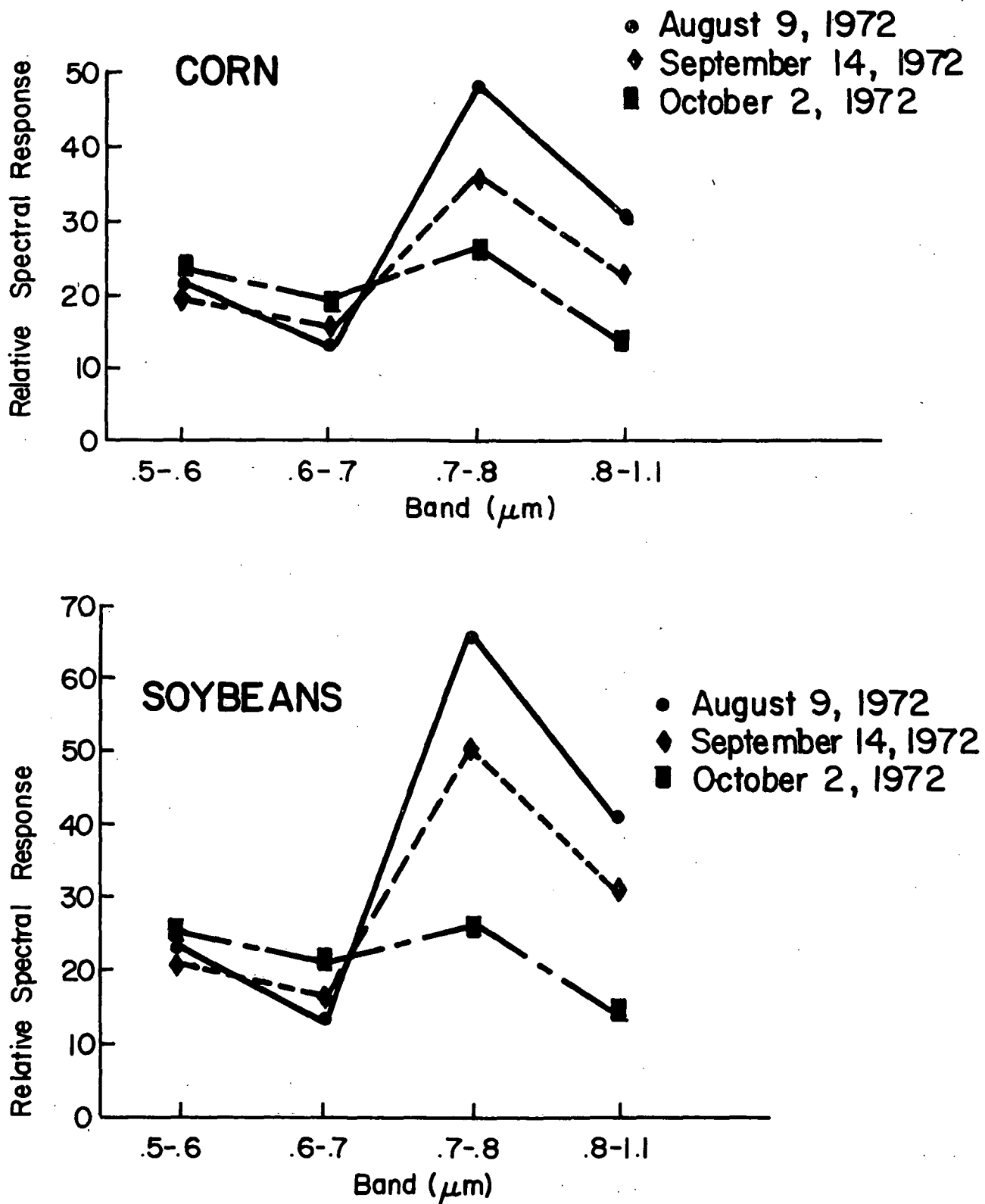


Figure 2.6. Mean relative spectral response of corn and soybeans in Northern Illinois on three dates.

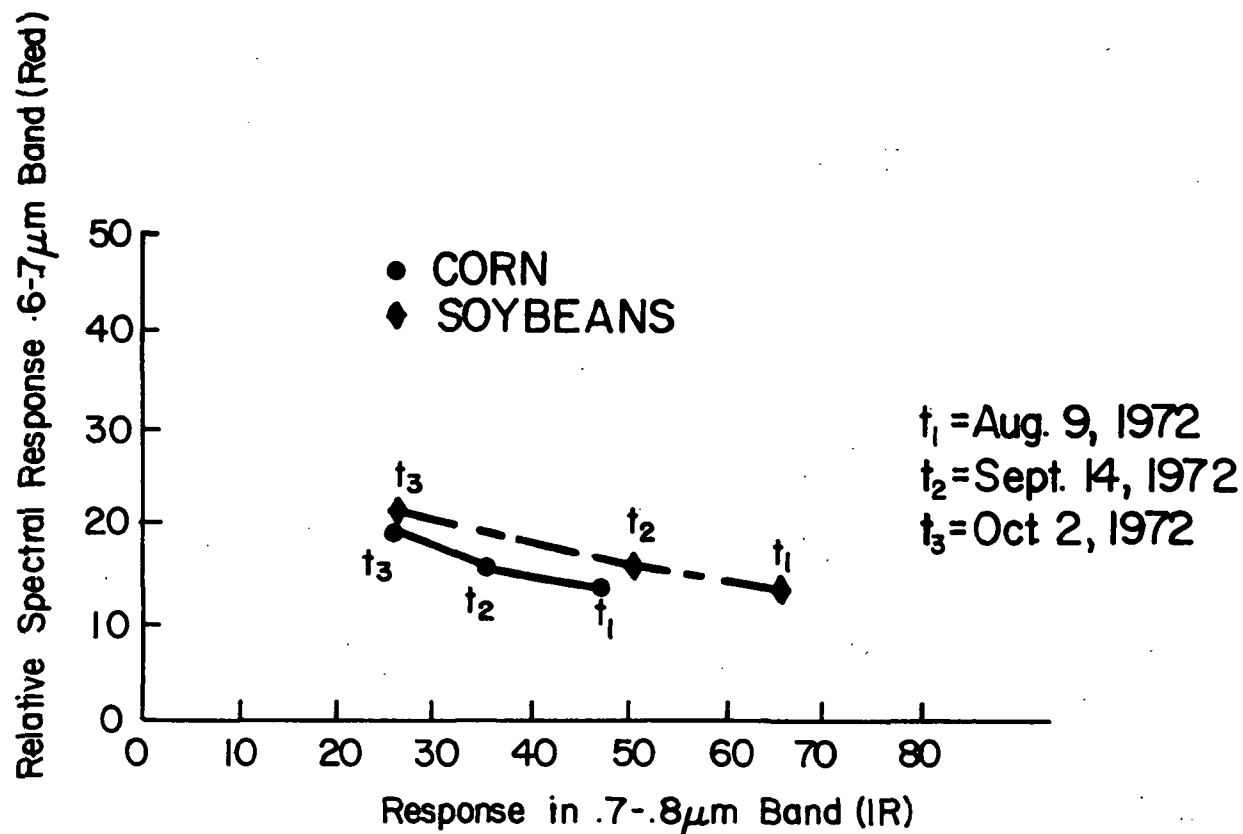


Figure 2.7. Mean ERTS spectral response values for two channels plotted as a function of time for fields of corn and soybeans in Northern Illinois, 1972.

were determined and a classification was performed using the same training and test fields as before. The feature selection processor chose two channels from August 9 and two channels from September 19. The results for this experiment are included in Table 2.15. The overall training field accuracy increased by 3.5% while the test field results decreased 1.5% compared to the four channel August 9 results. The greatest influence on this decrease was the corn test accuracy decrease from 76.3% to 72.5%. The best six of the twelve available channels produced an increase in training accuracy of 5.3% but the test accuracy still was reduced, 2.5% in this case. All 12 available channels were also used and training accuracy increased only 4.5%, less than for the six channel case, and test accuracy decreased 4.4%.

These results suggest that straight forward inclusion of spectral measurements offering generally less spectral separability with a near optimum set of measurements will not improve the overall results of the classification. A correct temporal classification algorithm would use any new information made available by the temporal dimension and would not decrease classification accuracy if no new information existed in the new (temporal) data as was the case in these experiments. Thus, further study is required to define classification procedures which will make optimum use of the spectral and temporal measurements available.

The value of temporal data for improving classification results with non-optimal data was also investigated. Assuming only September and October data were available, a classification analysis was performed using the best four of the September and October channels. Table 2.15 also contains the results of this experiment. The test field accuracy using only September channels was 58.1% and the best October results were 56.2%. Using the best four channels from September and October which includes the spectral and temporal dimensions, the test results were 67.8%.

Table 2.15 Comparison of classification results obtained from the spectral dimension and the spectral plus temporal dimensions of ERTS data.

| Class | Spectral | | | Spectral + Temporal | | | |
|-------|----------|----------|--------|--------------------------|--------------------------|----------------------|--------------------------|
| | Aug. 9 | Sept. 19 | Oct. 2 | Best 4 Channels, 3 Times | Best 6 Channels, 3 Times | 12 Channels, 3 Times | 4 Channels, Sept. & Oct. |

Training Field Classification (% Correct)

| | | | | | | | |
|----------|------|------|------|------|------|------|------|
| Corn | 87.3 | 82.2 | 79.8 | 92.9 | 94.6 | 92.7 | 92.0 |
| Soybeans | 90.6 | 62.5 | 71.4 | 96.0 | 96.0 | 96.0 | 86.7 |
| "Other" | 94.0 | 52.5 | 87.6 | 91.7 | 95.4 | 95.9 | 89.9 |
| Overall | 89.9 | 69.5 | 79.6 | 93.4 | 95.2 | 94.4 | 89.8 |

Test Field Classification (% Correct)

| | | | | | | | |
|----------|------|------|------|------|------|------|------|
| Corn | 76.3 | 58.9 | 55.3 | 72.5 | 73.0 | 68.4 | 64.6 |
| Soybeans | 84.9 | 58.1 | 49.2 | 82.2 | 83.5 | 81.0 | 76.1 |
| "Other" | 64.6 | 53.6 | 68.8 | 78.7 | 66.1 | 81.0 | 85.0 |
| Overall | 75.9 | 58.1 | 56.2 | 74.4 | 73.4 | 71.5 | 67.8 |

In this case, use of temporal data did have an appreciable beneficial effect on classification accuracy. Test results improved by 9.6% over the best of two purely spectral cases and were 11.5% better than the worst case. This result suggests that there is some "crossover" point at which temporal data begins to improve classification results rather than harm them with reference to purely spectral classifications using data obtained only at one time.

The overall conclusion drawn from this experiment is that use of the temporal dimension will require more complex procedures than now used with multispectral data and further research into methods for utilizing this new data dimension is required.

2.342 SPATIAL RESOLUTION AND FIELD SIZE

The 80 meter instantaneous field of view of the ERTS MSS data is considerably less than from aircraft scanner data previously analyzed for crop identification. This led to some concern about accurately locating crop fields in the data. This is particularly critical in the case of training fields from which the statistics on which the subsequent classifications will be based are obtained. If the training fields for a class contain points from other classes or if the response of single resolution elements is from more than one cover type, poor classification is likely to result. For this reason, only what was believed to be "pure" samples were selected from each training or test field. A later test to verify this was performed by training and testing the classifier after the sample of points had been reduced by one line and column on all sides to increase the probability that only pure samples had been selected. The percent correct classification of corn, soybeans, and "other" test fields was 81.2, 76.6, and 52.7, respectively, with overall performance of 79.0%. These classification results were compared to those previously obtained (Table 2.6). It was

concluded that there was no significant difference between the two results and therefore, that the original samples must have been accurately located. We also found no relationship between field size (number of points selected from a field) and classification accuracy.

Another question is what is the minimum field size which can be accurately located in the ERTS data? Our results show that square fields as small as 4 hectares (10 acres) could be found if the surrounding fields were substantially different in appearance. Fields of this size, however were not sampled because of uncertainty that only pure samples would be selected. Generally, fields had to be 8 hectares (20 acres) or more to be sampled. This size assumes square or nearly square field shapes. Narrow fields (300 meters or less) were not sampled regardless of their total area.

While a reasonably large percentage of the fields in the Corn Belt area are large enough to be accurately located and sampled in ERTS data of this resolution, there are many areas of the United States, as well as in other countries, where this would not be true. It is our belief that the 80 meter scanner IFOV may not permit reliable crop surveys to be made from ERTS data in those areas where small fields predominate. The problem is that as field size decreases the percentage of pixels which are mixtures of two or more crops or cover types increases rapidly. If too high a percentage of pixels contain mixtures, it will be difficult if not impossible to accurately identify individual crop species. The 40 meter instantaneous field of view anticipated for the Earth Observation Satellite (EOS) should substantially improve the capability to conduct crop surveys in those areas having small fields.

2.34 SUMMARY AND CONCLUSIONS

A three-county area of ERTS data in northern Illinois was classified into corn, soybeans, and "other". Recognition of test fields was about 80% which compared favorably with results previously obtained for aircraft scanner data. Acreage estimates for the 3200 square kilometer (2000 sq. miles) area agreed very well with those made by the USDA. The use of class weights, classification unbiasing procedures, and spatial information in the classification yielded improved results. The extendability, variability, and size of training sets required was also determined. Additionally, wavelength band selection, the spectral characteristics of major cover types, and the use of multitemporal ERTS data were studied.

2.4 INDIANA STUDY

A second study investigating the use of ERTS MSS data for identifying crops and estimating their acreages was conducted over a 12 county area in northwestern Indiana. Originally we had planned to classify the ERTS data for the entire state of Indiana; however cloud-free ERTS coverage was not collected over the entire state during either the 1972 or 1973 growing seasons.

2.41 OBJECTIVES

The primary objective of this analysis was to evaluate the quality of crop acreage estimates for county and crop reporting district size units based on classifications of ERTS data. A considerable amount of effort was spent in the Illinois study developing and testing alternate methods of classifying the ERTS data and converting the classification to acreage information. In this study we applied those techniques to a new set of conditions, i.e. ERTS data from another

time and location, and evaluated the quality of the acreage estimates.

2.42 PROCEDURES

Two ERTS frames, 1394-16035 and 1394-16042, collected August 21, 1973 over a 12 county area in northwestern and west central Indiana were used for analysis (Figures 2.8 and 2.9). Information describing the land use and 1973 corn and soybean acreages in the twelve counties is presented in Tables 2.16 and 2.17. Ground observation data consisting of the crop identification and location of approximately 100 fields in each of five segments were recorded on black and white aerial photography. The locations of the segments, approximately 200 square km in size, are shown in Figure 2.8. The crops and other cover types were identified as corn, soybeans, or "other". The "other" included hay, pasture, small grain stubble, woods, and urban areas.

The ERTS data was geometrically corrected as described in reference 10 to facilitate accurate location of the coordinates of training and test fields and county boundaries. A random selection of half of the fields for which the crop identification was known were used for training the classifier according to the procedure defined by Davis and Swain⁶. The remainder of the fields were used to test the accuracy of the classification.

The individual segments were then classified, using class weights based on county crop acreage estimates made the previous year, and test field performances obtained and evaluated. Next, each of the 12 counties were classified with the training statistics from the segment within or closest to the county. Porter County statistics were used to classify LaPorte, Porter, and Starke Counties; White was used for White, Pulaski and Jasper Counties; and the combined statistics for Tippecanoe and Benton were used for Tippecanoe, Benton, Fountain and Warren Counties. The number of pixels classified into each class (corn, soybeans,

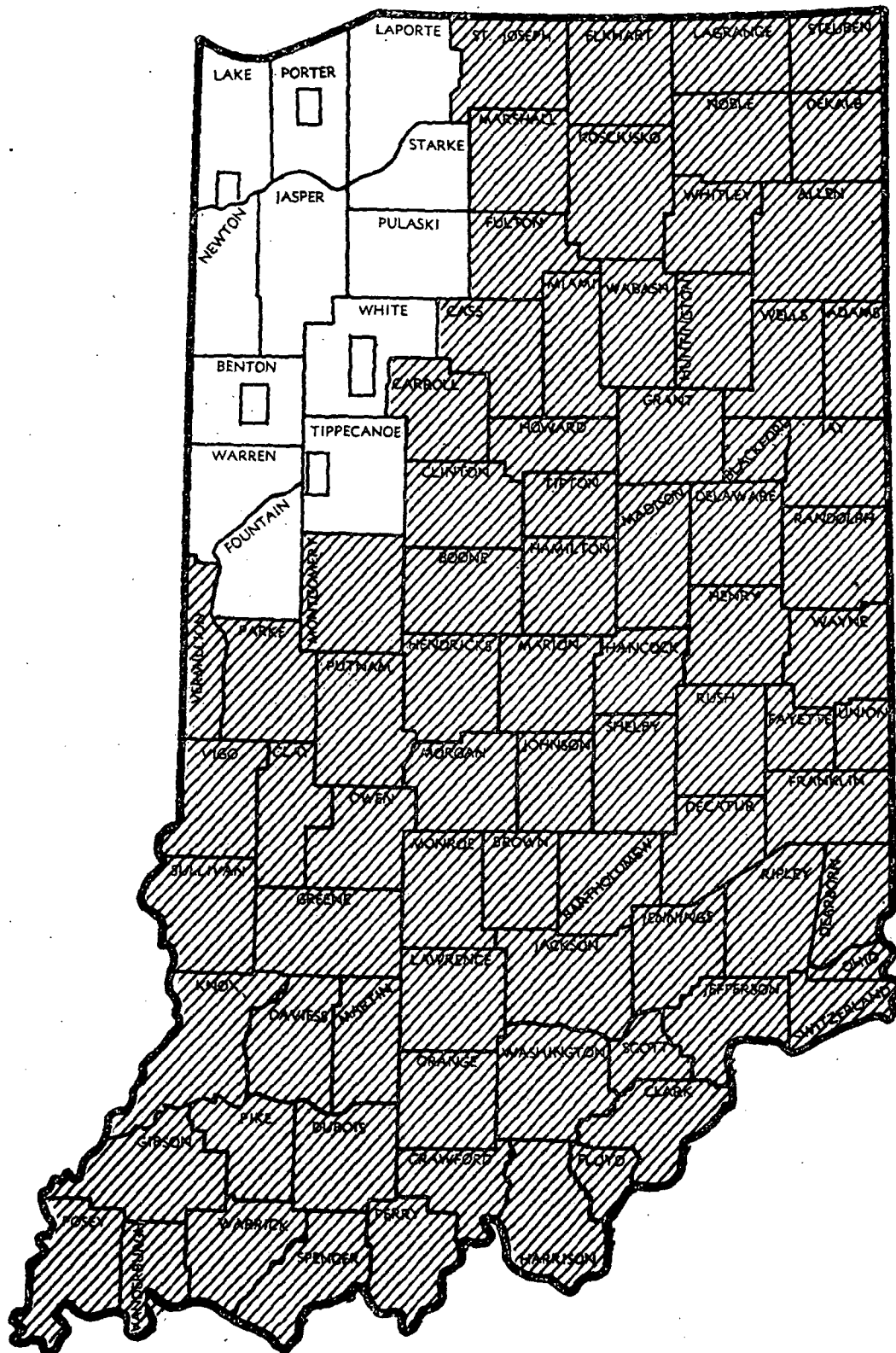


Figure 2.8 Indiana map showing 12-county test area.



Figure 2.9 ERTS imagery (band 5) of frames 1394-16035 and 1394-16042 collected August 21, 1973 over Indiana and Illinois.

Table 2.16 Summary of land use in northwestern Indiana counties, 1967.*

| County | Total Land Area | Non-Agricultural Land | | Agricultural Land | | | |
|----------------|-----------------------|--------------------------|-------------------------|-------------------|---------|--------|-------|
| | | Urban and Built-Up | Small Water Areas | Cropland | Pasture | Forest | Other |
| Thousand Acres | | | | | | | |
| Lake | 329.0 | 38.0 | 0.7 | 175.0 | 12.6 | 10.2 | 92.4 |
| Porter | 272.0 | 35.0 | 0.8 | 165.4 | 11.0 | 27.6 | 32.0 |
| LaPorte | 389.1 | 36.8 | 10.9 | 250.7 | 28.7 | 28.0 | 32.4 |
| Starke | 199.0 | 13.0 | 0.2 | 131.2 | 3.4 | 27.0 | 24.1 |
| Pulaski | 277.1 | 7.3 | 0 | 207.8 | 13.1 | 32.0 | 16.8 |
| Jasper | 359.1 | 10.1 | 0.1 | 272.5 | 34.4 | 25.6 | 16.4 |
| Newton | 262.8 | 5.4 | 0 | 188.8 | 25.2 | 19.2 | 23.5 |
| White | 318.0 | 9.6 | 0 | 269.0 | 18.7 | 12.8 | 7.9 |
| Benton | 261.7 | 6.2 | 0.1 | 240.1 | 7.2 | 2.0 | 6.0 |
| Warren | 235.5 | 5.5 | 0.9 | 164.9 | 29.9 | 23.3 | 10.9 |
| Fountain | 254.1 | 8.5 | 1.1 | 170.0 | 34.0 | 27.4 | 12.9 |
| Tippecanoe | 320.6 | 18.7 | 0.2 | 225.0 | 23.7 | 24.6 | 28.4 |
| TOTAL | 3,478.0 | 194.1 | 15.0 | 2,460.4 | 241.9 | 259.7 | 303.7 |

*Indiana Soil and Water Conservation Needs Inventory, 1967.

Table 2.17 Estimated acreages of corn, soybeans, and "other" in northwestern Indiana, 1973.*

| County | Total Land | Corn | Soybeans | "Other" |
|----------------|------------|---------|----------|---------|
| Thousand Acres | | | | |
| Lake | 329.0 | 62.2 | 45.4 | 221.4 |
| Porter | 272.0 | 63.5 | 43.4 | 165.1 |
| LaPorte | 389.1 | 108.7 | 64.6 | 215.8 |
| Starke | 199.0 | 69.9 | 35.5 | 93.6 |
| Pulaski | 277.1 | 96.3 | 83.5 | 97.3 |
| Jasper | 359.1 | 141.2 | 89.9 | 128.0 |
| Newton | 262.8 | 103.7 | 64.2 | 94.9 |
| White | 318.0 | 123.7 | 105.4 | 88.9 |
| Benton | 261.7 | 103.6 | 111.0 | 47.1 |
| Warren | 235.5 | 67.2 | 67.0 | 101.3 |
| Fountain | 254.1 | 71.7 | 65.4 | 117.0 |
| Tippecanoe | 320.6 | 96.2 | 81.7 | 142.7 |
| TOTAL | 3,478.0 | 1,107.2 | 857.0 | 1,513.1 |

*Annual Crop and Livestock Summary, 1973.

Table 2.18 Training and test field classification performances (% correct).

| County | Training Fields | | | Overall | Test Fields | | | Overall |
|------------|-----------------|----------|---------|---------|-------------|----------|---------|---------|
| | Corn | Soybeans | "Other" | | Corn | Soybeans | "Other" | |
| Lake | 66.7 | 79.5 | 78.0 | 74.7 | 59.3 | 80.0 | 35.2 | 57.5 |
| Porter | 88.0 | 69.9 | 51.5 | 79.0 | 85.1 | 78.1 | 46.7 | 79.4 |
| Tippecanoe | 78.4 | 93.9 | 80.4 | 83.7 | 64.2 | 86.5 | 82.0 | 76.0 |
| Benton | 74.4 | 91.5 | 85.3 | 83.3 | 81.1 | 82.4 | 78.8 | 80.8 |
| White | 86.5 | 95.9 | 62.3 | 85.6 | 74.3 | 71.1 | 47.9 | 68.0 |

and "other") was tabulated and compared to acreage estimates made by SRS/USDA. Standard statistical tests (paired-t test for comparison of sample means) were applied to test the hypothesis that the mean difference between the ERTS and USDA acreage estimates was not significantly different than zero.

2.43 RESULTS AND DISCUSSION

2.431 CLASSIFICATION OF TEST FIELDS

The results from this study consist of the test field classification performance and acreage estimates derived from the ERTS classifications. A summary of the test field classification performances is presented in Table 2.18. While the best performances are similar to those previously obtained in the Illinois study, there is considerably more variation among the segments in the classification performance. The Tippecanoe and Benton County segments where classification accuracy of test fields was about 80% are similar to the Illinois test site. The soils are uniform, fields are relatively large, and corn and soybeans are the major crops present in those segments.

The lower test field recognition for the Lake, Porter, and White County segments is attributed to the greater variability present in the ground scene than in the Illinois test site or the Benton and Tippecanoe County segments. For example, about half of the Lake County segment was river bottom land, where the crops had a distinctly different spectral response than the remainder of the segment. Similarly, the Porter County segment contained several major different soils which in turn had different cropping conditions even though corn and soybeans were still the primary crops present. With limited ground observation data available for training the classifier it was not possible to fully account for all the conditions present. With more cover types present the probability that one or more of them will be spectrally similar to corn and/or soybeans is increased.

The difficulties of attempting to extend training statistics over areas containing too much variability are indicated by the classification performances obtained when classifying one segment with statistics from another segment (Table 2.19). In general, performance decreased when "non-local" statistics were used to classify a segment. Such results are an indication that the two areas are not the same.

It appears then that within an area having relatively homogeneous soils, cover types, and cropping practices that the major crops can be accurately classified, but when other areas containing additional variability are included the recognition accuracy decreases. In other words, with more cover types and conditions present the probability that one or more of them will be spectrally similar to corn and/or soybeans is increased. To overcome this problem we recommend that the ERTS scene be divided or stratified into areas which are homogeneous prior to classification. We believe the ERTS imagery itself will be the best medium for delineating uniform areas, but other information such as soil association maps may also be useful. This approach is being tested as part of NASA Contract NAS 9-14016.

2.432 ACREAGE ESTIMATES

The number of pixels classified into each class for the individual counties are shown in Table 2.20. Corn and soybean acreage estimates made by the USDA in 1971 for each county were used for each class weights. The unbiasing procedure used in Illinois was not applied to the classifications because of concern that the estimates of error rates were not accurate due to the small number of test fields available and that the segments containing the test fields were not representative of the entire counties.

Acreage estimates based on the ERTS classification are compared to the county crop acreage estimates made by the USDA in Table 2.21. The agreement between the two estimates ranges from

Table 2.19 Comparison of test field classification accuracies obtained with local and non-local training statistics.*

| Segment Classified | Statistics From | Percent Correct Classification of | | | |
|--------------------|-----------------|-----------------------------------|----------|---------|---------|
| | | Corn | Soybeans | "Other" | Overall |
| Lake | Lake | 48.0 | 79.1 | 43.1 | 53.1 |
| | Porter | 50.1 | 67.8 | 73.7 | 60.3 |
| Porter | Porter | 69.3 | 66.2 | 79.1 | 72.7 |
| | Lake | 67.4 | 80.8 | 62.2 | 68.4 |
| Benton | Benton | 64.2 | 86.5 | 82.0 | 82.0 |
| | Tippecanoe | 69.4 | 88.3 | 89.0 | 80.2 |
| Tippecanoe | Tippecanoe | 64.2 | 79.1 | 74.4 | 72.0 |
| | Benton | 78.7 | 74.8 | 71.2 | 75.8 |
| White | White | 74.3 | 71.1 | 47.9 | 68.8 |
| | Benton | 63.9 | 57.7 | 75.8 | 63.1 |
| | Tippecanoe | 70.0 | 61.1 | 64.2 | 65.3 |

*Equal class weights used in the classifications.

Table 2.20 ERTS classification results for 12 northwestern and west central Indiana counties.

| County | Total Points | No. Points Classified As | | |
|------------|-----------------|--------------------------|----------|---------|
| | | Corn | Soybeans | "Other" |
| Lake | 199,777 | 24,366 | 41,118 | 134,293 |
| Porter | 167,932 | 32,432 | 16,155 | 119,345 |
| LaPorte | 238,387 | 57,387 | 26,924 | 154,076 |
| Starke | 110,807 | 31,723 | 16,301 | 62,783 |
| Pulaski | 92,335 | 32,031 | 24,882 | 35,422 |
| Jasper | 231,921 | 90,057 | 60,844 | 81,020 |
| Newton | 159,177 | 53,670 | 56,048 | 49,459 |
| White | 176,259 | 62,958 | 56,886 | 56,415 |
| Benton | 167,504 | 63,420 | 52,730 | 51,354 |
| Warren | 147,620 | 51,225 | 40,542 | 55,853 |
| Fountain | 162,109 | 52,400 | 42,015 | 67,694 |
| Tippecanoe | 119,775 | 42,170 | 28,930 | 48,675 |
| TOTAL | 1,973,603 | 593,839 | 463,375 | 916,389 |

Table 2.21 Comparison of USDA and ERTS acreage estimates
(Percent of total land area) for 12 Indiana
counties.

| County | Corn | | Soybeans | | "Other" | |
|------------|------|------|----------|------|---------|------|
| | USDA | ERTS | USDA | ERTS | USDA | ERTS |
| Lake | 18.9 | 12.2 | 13.8 | 20.6 | 67.3 | 67.2 |
| Porter | 23.3 | 19.3 | 16.0 | 9.6 | 60.7 | 71.1 |
| LaPorte | 27.9 | 24.1 | 16.6 | 11.3 | 55.5 | 64.6 |
| Starke | 35.1 | 28.6 | 17.8 | 14.7 | 47.0 | 51.7 |
| Pulaski | 34.8 | 34.7 | 30.1 | 26.9 | 35.1 | 38.4 |
| Jasper | 39.3 | 38.8 | 25.0 | 26.2 | 35.6 | 34.9 |
| Newton | 39.5 | 33.7 | 24.4 | 35.2 | 36.0 | 31.1 |
| White | 38.9 | 35.7 | 33.1 | 32.3 | 28.0 | 32.0 |
| Benton | 39.6 | 37.9 | 42.4 | 31.5 | 18.0 | 30.6 |
| Warren | 28.5 | 34.7 | 28.5 | 27.5 | 43.0 | 32.8 |
| Fountain | 28.2 | 32.3 | 25.7 | 25.9 | 46.0 | 41.8 |
| Tippecanoe | 30.0 | 35.2 | 25.5 | 24.2 | 44.5 | 40.6 |
| Total | 35.1 | 30.1 | 29.2 | 23.5 | 35.7 | 46.4 |

excellent (Jasper Co.) to poor (Lake Co.). Most counties, however, fall between these extremes. Disagreement between the two estimates may be due to errors in either or both estimates; however, for this study we are using the USDA estimates as the standard or reference for evaluating the estimates. The best ERTS estimates were for those counties classified with statistics from the segments in Benton, Tippecanoe, and White Counties. All of these counties tend to have similar soils and cropping practices. And, the predominant land use is agricultural crops. Lake, Porter, and LaPorte Counties, on the other hand, have large urban areas and a wider range of cover types which were not well represented in the training. For instance, the northern third of LaPorte County has a large expanse of rolling land covered by orchards. Since no orchards were included in statistics used to classify LaPorte County, some misclassification can be expected there.

The average differences between the USDA and ERTS estimates were 1.7, 1.4, and 3.1 percent for corn, soybeans, and "other", respectively. These differences were not statistically significant as indicated by paired-t tests. However, the USDA and ERTS estimates of the total acreage of each crop present in the 12-county area were substantially different.

There are several possible causes of poor agreement between the two estimates: (1) Errors may exist in either USDA or ERTS estimates; however, the USDA is probably within 10 percent of the true acreage. (2) The USDA estimate for corn is the acreage harvested for grain, whereas the ERTS classification is of all corn; however, this difference would be less than 5% of the corn acreage and does not account for all the difference. (3) The entire county was not included for Starke, Pulaski, White, and Tippecanoe Counties in the ERTS data. Although for this study we have assumed that the percentages of each crop are reasonably constant across an entire county, there undoubtedly are differences. But, again this does not explain all the differences

between the ERTS and USDA estimates, e.g. Porter County which was included in its entirety.

2.433 SUMMARY AND CONCLUSIONS

While the results from this study are not as positive as those from the Illinois study, a considerable amount of knowledge which can be utilized in the future to make improvements was gained. The results do point out the difficulties in classifying large areas containing numerous cover types. Several suggestions on how to alleviate problems encountered in this analysis were made. These include a recommendation to have sufficient ground observation data available to characterize all of the cover types or conditions present in the ERTS scene and to stratify the scene into relatively homogeneous areas prior to classification.

2.5 MISSOURI AND IDAHO STUDIES

In addition to the investigations in Illinois and Indiana LARS assisted the Research and Development Branch, Statistical Reporting Service, U.S. Department of Agriculture (SRS/USDA) in the analysis of ERTS data collected over Missouri and Idaho. This work had the general objectives of broadening our experience in the analysis of ERTS data from areas having different crops and conditions than found in the Corn Belt and increasing the capabilities of SRS/USDA to analyze and evaluate remote sensing data. Since they will be fully reported in the final report of SRS's ERTS investigation the approach and results will be described only briefly in this report.

2.51 ANALYSIS OF MISSOURI DATA

An analysis of ERTS data collected during the summer and early fall, 1972, over Crop Reporting District No. 9 in Southeast Missouri was conducted. SRS supplied the ground truth data and assisted in the analysis of the MSS data, LARS geome-

trically corrected and overlaid the ERTS MSS data, located the ground truth segments and fields in the data and worked with SRS in analyzing the MSS data.

Twenty-nine area segments were located in two ERTS frames covering Crop Reporting District No. 9 in Southeast Missouri. Data from ERTS passes on August 26, September 14, and October 2, 1972 were overlaid and geometrically corrected. Geometric correction greatly facilitated locating segments and fields. Temporal overlay alleviated the necessity of locating fields in three different data sets as well as permitted a test of the usefulness of temporal data in the classification.

Segments were located in the August ERTS data which had been deskewed and scaled to 1:24,000 scale by overlaying computer printouts onto 1:24,000 scale maps on which the segments had been drawn. The segments were then clustered and coordinates of the individual fields found on a non-supervised classification map. Statistics for the classes of cotton, soybeans, corn, harvested wheat, grass, and miscellaneous were obtained and the data classified. Nearly all the available crop fields were used as training fields.

Results of the classifications are shown in Table 2.22. These are for the multitemporal case where bands from three ERTS passes were used. While reasonably good classification of cotton and soybeans was achieved, identification of "other" even when considered as one class was poor. It should be noted that these results are of training fields for multitemporal data. Test field performance is generally somewhat lower.

The value of multitemporal information in the classification of cotton and soybeans is shown in Table 2.23. For cotton, performance was improved 7 to 19% by using all bands from three dates compared to each of the dates individually. For soybeans, the highest performance was for the single August 26 ERTS pass.

Table 2.22 Training field classification performance for 29 segments of ERTS data from southeastern Missouri.

| Class | No. Points | No. Points Classified As | | | Percent Correct |
|----------|------------|--------------------------|------------|------------|-----------------|
| | | Cotton | Soybeans | "Other" | |
| Cotton | 927 | 739 | 137 | 51 | 79.7 |
| Soybeans | 852 | 99 | 612 | 141 | 71.8 |
| "Other" | <u>438</u> | <u>68</u> | <u>117</u> | <u>253</u> | <u>57.8</u> |
| TOTAL | 2,217 | 906 | 866 | 445 | 72.4 |

Table 2.23 Comparison of single date and multitemporal classification of cotton and soybeans.

| Crop | Classification Performance (Percent Correct) | | | |
|----------|--|---------------------------|------------------------|-----------|
| | August 26 ¹ | September 14 ² | October 2 ³ | All Dates |
| Cotton | 60.6 | 69.7 | 73.2 | 79.7 |
| Soybeans | 86.0 | 67.6 | 62.4 | 71.8 |

- ¹ Bands 4, 5, and 7
² Bands 5 and 7
³ Bands 4, 5, 6, and 7

2.52 ANALYSIS OF IDAHO DATA

An analysis of ERTS data for crop species identification was also conducted in southeastern Idaho over a crop reporting district approximately the size of the ERTS frame. The ERTS data, scene 1035-17525, was collected August 27, 1972. Procedures similar to those used in Missouri were used in the classification. However, with 65 segments averaging more than four square kilometers in size there were considerably more fields available for training and testing the classifier. Clustering was used to define 24 subclasses of the 10 classes present.

The Idaho test site was a diverse agricultural area with a wide range of crops including corn, alfalfa, sugar beets, potatoes, beans, harvested winter wheat, barley, spring wheat, pasture, and fallow land. Classification results for test fields containing a total of 7271 points are shown in Table 2.24. Overall performance was only about 40% for the 10 classes. While this performance is very low it should be noted that we were attempting to identify more classes (10) over a larger area (an entire ERTS frame) than in the previous classifications.

The greatest source of confusion among the classes was pasture. Almost twice as many points were classified as pasture than were actually present. There was also considerable confusion among the other classes, too, as indicated by the classification accuracies ranging from 4 to 64%.

2.53 SUMMARY AND CONCLUSIONS

Two experiments were conducted over geographic areas of 4,000 to 8,000 square kilometers. In each case the recognition of test fields was less than the 80% accuracies previously reported by LARS and other investigators. There are major differences, however, in the size of areas being classified and the diversity and complexity of the ground scene of the Missouri and Idaho test sites compared to Illinois. Our conclusions from these analyses are that the more classes (crops)

Table 2.24 Classification performance of test fields in
southeastern Idaho.

| Crop | No. Points | % Correctly Classified |
|-----------------|------------|------------------------|
| Alfalfa | 1,314 | 29.8 |
| Pasture | 1,433 | 64.0 |
| Barley | 957 | 25.9 |
| Harvested Wheat | 813 | 62.6 |
| Spring Wheat | 104 | 3.8 |
| Corn | 541 | 8.5 |
| Beans & Peas | 549 | 40.6 |
| Sugarbeets | 386 | 56.0 |
| Potatoes | 395 | 21.8 |
| Fallow | <u>779</u> | <u>37.4</u> |
| OVERALL | 7,271 | 40.3 |

there are to identify the greater the probability of misclassification. This is particularly true when all or most of the classes consist of green vegetation. More wavelength bands in the middle and far infrared would probably improve the capability to make the more subtle distinctions required to separate individual species.

As indicated in the Indiana study, working over a large area containing much variation in the crops and soils appears to lead to poor classification performance when the same procedures as developed for smaller homogeneous areas are used. A simple solution to this problem may be to divide the region into smaller, more uniform areas prior to classification. A more sophisticated approach would be to develop classifiers capable of adapting to a changing ground scene.

2.6 CONCLUSIONS FROM CROP IDENTIFICATION STUDIES

The overall conclusions from the crop identification and acreage estimation phase of this investigation are that the combination of ERTS-1 MSS data and machine processing of it can be used to obtain crop production information over large areas of the world. It has been shown in this study that it is possible to accurately identify major crop species from ERTS data and to convert the identification data to accurate estimates of the crop acreage using machine processing methods. The best performance is obtained when the data is collected at the right time in the crop's growth cycle, the fields are relatively large and uniform, and there are not too many crops to be identified. On the other hand, if there are several crops having similar characteristics or if the area to be classified is heterogeneous in its composition of crops and condition, ERTS data may not have sufficient spectral bands and dynamic range to enable accurate identification of individual species.

More specifically the following conclusions can be drawn from these studies:

1. An earth orbiting satellite is an excellent vehicle for rapidly collecting MSS data over large areas. Such a capability is particularly important for agricultural crops because crop production is dynamic and because information is needed on a world-wide basis.

2. Machine processing methods utilizing pattern recognition techniques such as LARSYS are well-suited to analyzing the large amounts of data collected by the ERTS-1 MSS and converting the data to useful information. In our experience it has not been possible to obtain quantitative information on crops from ERTS imagery by standard interpretation techniques.

3. The quality of the ERTS-1 MSS data is good; however, the 80 meter instantaneous field of view is a limitation in areas having small fields and the four bands are a minimum for producing accurate classifications. Additional wavelength bands in the middle and thermal infrared would undoubtedly improve classification performance, particularly in those areas having more than two or three major crops to be identified.

4. Although the spatial and spectral resolution of the ERTS-1 MSS data is limited compared to aircraft scanners, it does have several important advantages over aircraft data. These include: (1) a narrow field of view so that view angle is not a factor in the scene and (2) constant solar elevation at the time of data collection over large areas so that changing illumination conditions are less of a problem in identifying crops.

5. Cloud cover may be a limiting factor in some instances, but in an operational environment where data were being analyzed from over large areas (rather than small pre-designated test sites) this might very well be a less serious problem.

Still, in some agricultural situations more frequent collection would add to the value of the data.

6. While ERTS-1 has the capability to collect data over large areas, some care in analyzing it is required. In particular, because data from a large area is readily available, there may be some tendency to extend training statistics farther than the changing ground scene permits. The actual distance that statistics can be successfully extended before performance deteriorates will depend on how much and how fast the composition of the ground scene (i.e. the crops and other surrounding cover types) change. Stratification of large areas into smaller, more uniform areas prior to classification should help to alleviate this problem.

7. The use of the spatial and temporal dimensions of ERTS data can be expected to improve classification performance. However, data analysis techniques for utilizing these dimensions in addition to the spectral dimension must be developed before the full potentials of the spatial and temporal dimensions are reached.

8. It will continue to be important to have data from the right time of the year in terms of the crops being identified, e.g. it will not be possible to identify all crops at any time during their growth cycle. In the end, of course, when suitable times are depends not only on the crop(s) of interest but also on the surrounding cover types as well. In this regard it will be important that the data analysts be familiar with the crops and area being classified.

9. The analysis of ERTS data is handicapped by the six to eight week interval between data collection and receipt of the data tapes and imagery. This is a particularly serious problem for agricultural crops which change quite rapidly and may even be harvested before the data is received. In order to carry out the best analysis, ground observation data needs to be collected very near the time of ERTS data collection. However, analysts are reluctant to spend a lot of time and effort

collecting ground truth until they know that cloud-free ERTS data was collected. They are, therefore, faced with the choice of collecting ground observation data which may never be used or trying to collect the necessary data after it may be too late. Neither alternative allows for optimum use of the ERTS data.

2.7 REFERENCES, SECTION 2

1. Eisgruber, L. M. 1972. Potential Benefits of Remote Sensing: Theoretical Framework and an Empirical Estimate. Information Note 030872. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.
2. Ewart, N. 1972. The Impact of Commodity Pricing--A Simulation Study. Ph.D. Thesis, Purdue University, West Lafayette, Indiana.
3. MacDonald, R. B., M. E. Bauer, R. D. Allen, J. W. Clifton, J. D. Erickson, and D. A. Landgrebe. 1972. Results of the 1971 Corn Blight Watch Experiment. Proc. of Eighth Int'l. Symp. on Remote Sensing of Environment, Ann Arbor, Michigan. October 2-6, 1972.
4. Swain, P. H. 1972. Pattern Recognition: A Basis for Remote Sensing Data Analysis. Information Note 111572. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.
5. Kumar, R. and L. F. Silva. 1974. Statistical Separability of Agricultural Cover Types in Subsets of One to Twelve Spectral Channels. Proceedings of Ninth Int'l. Symp. on Remote Sensing of Environment, Ann Arbor, Michigan. April 15-19, 1974.
6. Davis, B. J. and P. H. Swain. 1974. An Automated and Repeatable Data Analysis Procedure for Remote Sensing Applications. Proc. of Ninth Int'l. Symp. on Remote Sensing of Environment, Ann Arbor, Michigan. April 15-19, 1974.
7. Gupta, J. N., R. L. Kettig, D. A. Landgrebe, and P. A. Wintz. 1973. Machine Boundary Finding and Sample Classification of Remotely Sensed Agricultural Data. Proc. of Conf. on Machine Processing of Remotely Sensed Data, West Lafayette, Indiana. October 16-18, 1973.
8. Steiner, D. 1970. Time Dimension for Crop Surveys from Space. Photogrammetric Engineering, 36:187-194.
9. Anuta, P. E. 1970. Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques. IEEE Transactions on Geoscience Electronics, GE-8:353-368.
10. Anuta, P. E. 1973. Geometric Correction of ERTS-1 Digital Multispectral Scanner Data. Information Note 103073. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

3.0 Soil Association Mapping

3.1 Introduction

Computerized analysis of ERTS MSS data has yielded images which will prove useful in the ongoing Cooperative Soil Survey program, involving the Soil Conservation Service of USDA and other state and local agencies. In the present mode of operation, a soil survey for a county may take up to 5 years to be completed. Results reported here indicate that a great deal of soils information can be extracted from ERTS data by computer analysis. This information is expected to be very valuable in the premapping conference phase of a soil survey for a county, resulting in more efficient field operations during the actual mapping. It is expected to result in greater accuracy of mapping and decrease the time required to produce the soil survey.

The work reported here is concerned primarily with comparison of generalized county soil maps with multispectral maps produced by computer analysis of ERTS MSS data. Initial investigation of discriminability of individual soil types for more detailed mapping was also conducted.

Results are reported for studies in Tippecanoe and White Counties in Indiana and in Finney County, Kansas. Early results have been reported previously and are not discussed further in this report. The procedures used and a more detailed evaluation of the results are given for the Tippecanoe County test site.

The Kansas test site was introduced after finding a strong relationship between soils maps and ERTS spectral images in Indiana. Many persons in the remote sensing community and soil surveyors had expressed concern that the Indiana results might not be typical of what could be expected to be attained with ERTS images. That is, they felt the Indiana results were too optimistic, and results would not be as encouraging in areas like Kansas where the soils are less contrasting.

3.2 White County, Indiana Analysis

Figure 3.1 is a geometrically corrected image of White County. The original was a simulated color infrared (false color) image produced from ERTS data using a digital image display system by exposing color film to the green band using a blue filter, the red band using a green filter, and the infrared band using a red filter. This was printed at a scale of 1:160,000. Red colors in the original image represented green vegetation while all other tones represented bare soils, non-green vegetation, or other scene features. Very little, if any, non-green vegetation was present in White County on June 9, 1973 when this data was collected, so it can be assumed that all tones other than the red represent bare soil and other features. In general, field observations have shown that the bright regions in this figure represent soils of rolling hills while the darker regions depict depressional areas having more nearly level topography.

The soil association boundaries in Figure 3.1 are from an existing generalized soil map, published in 1971. A properly scaled overlay of this soil map was made to fit the ERTS imagery. Comparing the existing soil association boundaries with the imagery shows that relationships between the boundaries and various tonal patterns in the imagery do exist. There are, however, distinct soil patterns in the image which have not been mapped on the general soil map.

3.21 Soil Association Discussion

Soil association 71, identified as the Randolph-Millsdale association on the soil association map, is actually comprised of Rensselaer, Darroch, and Granby soils. This incorrect identification is clearly evident from the imagery, because of the propoerties of these soils. (Rensselaer, Darroch and Granby

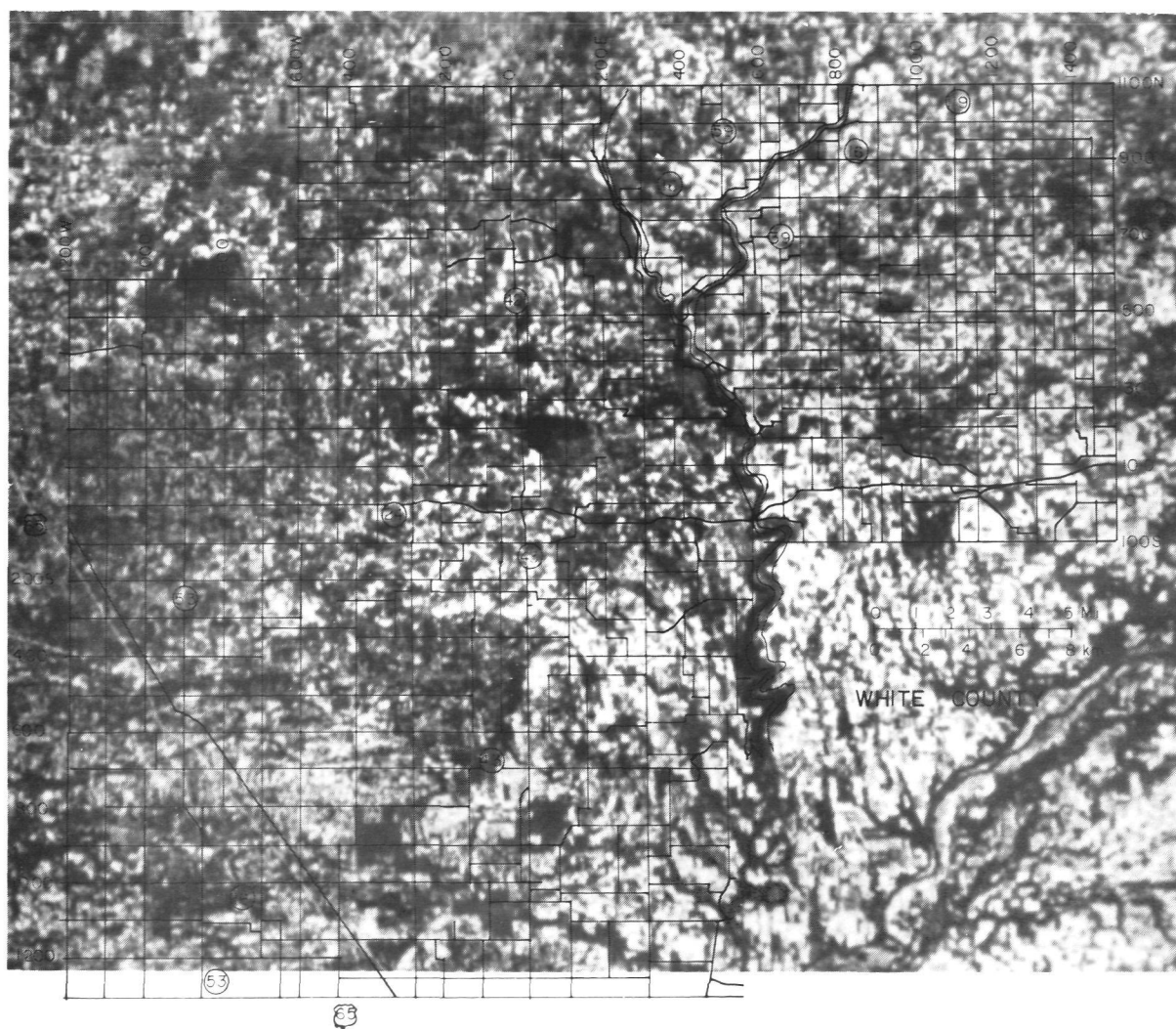


Figure 3.1 Black and white reproduction of simulated color IR from data obtained by ERTS June 9, 1973 over White County, Indiana. The solid lines represent soil association boundaries, while the dotted lines are areas of particular interest.

are lake plain soils.) The Rensselaer and Darroch are soils having loamy surfaces and are very poorly drained and somewhat poorly drained respectively. The Granby soils are sandy but are also very poorly drained. Because of these properties, i.e., being lake plain soils, and being poorly drained, they appear much the same as other glacial lake basin soils such as in association 23 and area A in the figure. Because these kinds of soils are poorly drained, their utility as residential areas is limited. However, they make productive agricultural lands when properly managed. These factors make it very important to be able to accurately map these soils.

In association 88 (Odell-Chalmers) there appears a pattern labeled A near the middle of the county which is not typical of the rest of the association. It has been discovered through field examination of this area that it is, in fact, predominantly very poorly drained Rensselaer soils and should have been mapped separately from association 88.

Area B in the figure is another example of improper mapping. It should have been included as part of association 70 (Parr - Corwin). Parr is the well-drained member of the catena while Corwin is the moderately well drained member. This soil association consists of gently sloping topography on glacial till plains and low moraines. These soils give high yields of corn, oats, soybeans, alfalfa and wheat when commercial fertilizers are applied because of their drainage properties. At the time of the original mapping it was difficult to place the boundary of this association in its proper place. Now, with the use of ERTS imagery such discrepancies become readily apparent.

Association 39 (Plainfield-Brems-Morocco) should be extended to include area marked C in the figure. This association consists of sloping sandy soils with Plainfield being excessively drained, Brems being moderately well drained, and Morocco

somewhat poorly drained. It is important to be able to map these soils accurately because of their limitations in agriculture. Since permeability is relatively rapid in these sandy soils, it is difficult to grow crops in them unless they are irrigated, and then it is limited to growing rye, wheat, or soybeans. A relatively large acreage of these soils is in low grade pasture land.

Association 64 (Crosby-Brookston) next to area A should be extended from the finger that extends in a westerly direction to include the bright area up to the boundary of association 39, just north of area A. A dashed line shows the extension as it should have been mapped. Crosby is a somewhat poorly drained soil, and Brookston is a very poorly drained soil. Runoff is slow and permeability is slow, and a large portion of this area is under cultivation. The principal crops are corn, soybeans and small grains.

Association 23 (Maumee-Gilford-Rensselaer) at area D in the northeast corner of the county should, in fact, be the very dark area approximately 1.25 km to the east. This dark area is typical of old lake plain soils, which is in general what association 23 represents. This means the soil was mapped correctly, and the discrepancy between the map and the ERTS image is due to preparation of the two for overlay. When this area (area D) is properly overlaid, the east county boundary line is located correctly as verified by correlating aerial photography, ground information and the ERTS image. The error in this case resulted from photographic distortions. The geometrically corrected data from the line printer was checked and this distortion was not present.

It is suggested that the ERTS image, if it had been available, would have helped the soil surveyor to more accurately place many of his soil boundary lines while mapping in the field.

Areas A, B, and C as previously mentioned are examples of this. The synoptic view from ERTS makes soil and landscape patterns over a large area more obvious than do aerial photos which soil surveyors have available.

3.22 Use of ERTS Imagery in Field Mapping

Overlaying a county road map of appropriate scale onto the ERTS imagery allows the soil surveyor to more easily locate himself in the field in relation to the imagery, as shown in Figure 3.2. Driving north on State Road 43 just outside of the town of Chalmers (located at county road 600S and State Road 43) the area labeled "A" in Figure 3.1 is apparent. It is extremely important in using ERTS images in the soil survey program to be able to pinpoint precise ground locations to make field observations and collect soil samples for laboratory analyses.

Computer printout maps of an area are also valuable in the field, especially with roads and other landmarks located properly on them. Because of their physical size, these maps are produced for 4 to 20 square mile sub-areas of the county for use in the field. A computer printout map of scale 1:20,000 can give some information on individual mapping units for a much smaller area. In ongoing investigations at LARS newly mapped detailed soil boundaries and 1:20,000 computer maps are being overlaid and compared. This scale of 1:20,000 is used because it matches the scale of the black and white aerial photography the soil surveyor is using as a base map. It appears that these computer maps and the 1:160,000 scale ERTS imagery will provide much useful information for the soil mappers who are producing detailed as well as general soil maps from field surveys.



Figure 3.2 Black and white reproduction of simulated color IR from data obtained by ERTS June 9, 1973 over White County, Indiana with a road map overlay.

3.23 Conclusions

The ERTS false color imagery makes it possible to distinguish between surface soils which have dissimilar properties. Overlaying pre-existing general soil boundaries with the geometrically corrected ERTS data makes it easy to compare the two. Such comparisons indicate that many soils could have been mapped more correctly using ERTS imagery as a base photo during the generalized soil mapping process.

3.3 Tippecanoe County, Indiana

At the time of the ERTS pass on June 9, 1973 field observations were made to locate agricultural fields in which the soil was recently cultivated, so that little or no vegetation was present. These areas were to be used later for characterizing soil signatures during computer analysis. It was determined from ERTS imagery, that about 55% of the county was non-vegetated on June 9.

3.31 Procedures

The data collected by ERTS were geometrically corrected in all four bands which were used in subsequent computer analysis. A false color image (Figure 3.3) was produced using the digital image display as previously described. The ERTS image of Tippecanoe County was then displayed on the video screen (digital display unit) and fields for which ground information had been collected were located. All of these fields of non-vegetated soil which were of sufficient size to accurately locate in the ERTS image were selected to use in the computer analysis procedure. Samples were also selected to characterize spectral signatures of various vegetation types and water.

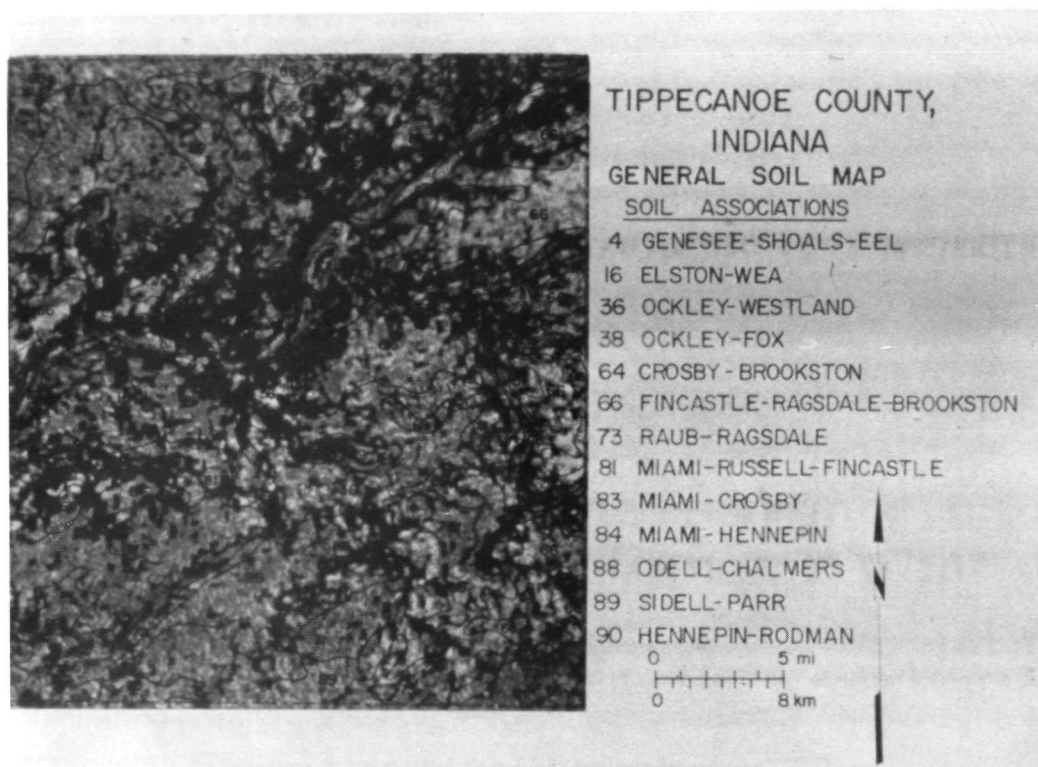


Figure 3.3 Black & White reproduction of a simulated color IR photograph from data taken June 9, 1973 over Tippecanoe County, Indiana; soil association map has been overlaid.

A multivariate cluster algorithm was used to determine the number of spectral classes of non-vegetated soils present in the data set. The definition of the spectral classes to be used was accomplished by partitioning the data set into decreasing numbers of spectral classes. A maximum of 20 and a minimum of 5 spectral classes were considered for data partitioning. An average separability of similar spectral classes was then computed. The measure of separability between classes was a multivariate distance (a "quotient"). This quotient computation incorporated covariance information. It has been noted from previous studies that the distribution of spectral classes in multidimensional space was often such that certain pairs of classes were similar to one another. Because this anomaly is usually observable in multispectral data processed by this analysis procedure, it was decided that spectral separability of similar classes only would be considered in determining an optimum number of classes. To accomplish this the average "quotient" was computed for only those classes which were spectrally most similar to one another.

The county was classified using a maximum likelihood algorithm into 10 spectral classes of soil, four spectral classes of vegetation, and two spectral classes of water. Distribution of the 10 spectral classes of soil was further examined. It was found that three of the 10 spectral classes were scattered randomly across the county and did not appear to relate to any particular soil or landscape conditions. They were therefore deleted from further analysis. Classification results with each of the spectral classes of soil color coded are shown in Figure 3.4. Soil association boundaries were manually overlaid directly onto the ERTS image without consideration of the spectral properties of the soils. Some colors are easily distinguishable, while others are less distinct, depending on the relative homogeneity of spectral classes.

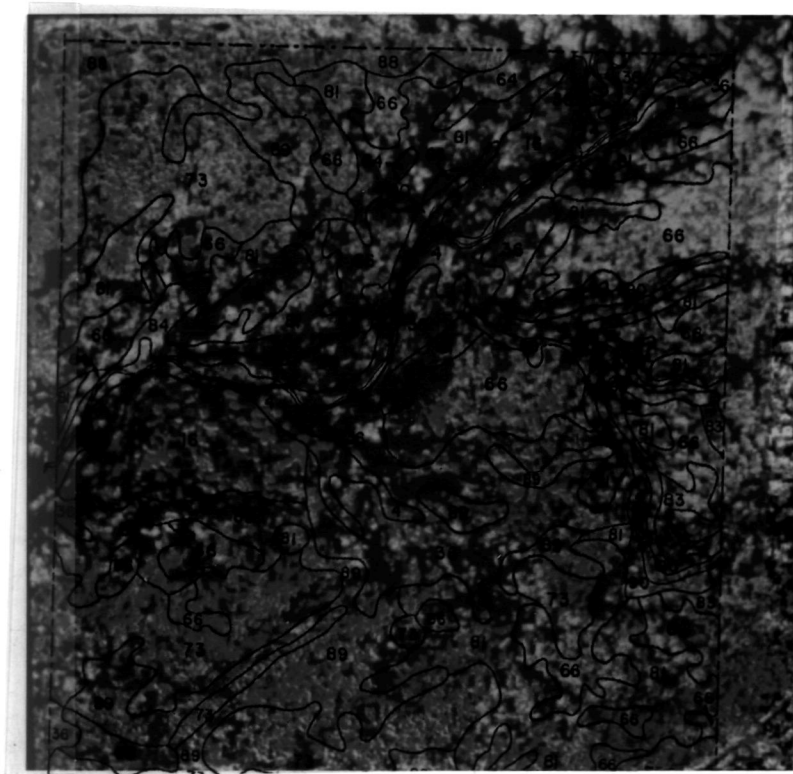


Figure 3.4 A classification result of Tippecanoe County data taken June 9, 1973.

3.32 Results

Figure 3.5 shows the spectral means and variances of the seven spectral classes of soil which were determined from the computer analysis. All seven classes were spectrally distinct from one another according to the previously defined criteria. Examination of the geographic occurrence of the seven spectral classes in Figure 3.4 permits some generalizations. Spectral classes one and two (Figure 3.4) are found mainly in the east half of the county in areas mapped as the Fincastle-Ragsdale-Brookston association (66) and the Miami-Russell-Fincastle association (81). In addition to being found predominantly in the eastern half of the county, these two spectral classes were often associated with relatively steep slopes or rolling topography. Spectral class three generally occurs in near proximity to spectral classes one and two. The soils found associated with these three spectral classes are predominantly well drained and somewhat poorly drained soils formed under forest vegetation. Typically these are the Miami, Russell, Fincastle, and Reeseville soils. They have predominantly silt loam surface textures.

Spectral class four, which is intermediate in reflectance in all four MSS bands, occurred primarily in an area mapped largely as the Ockley-Westland association (36), likewise an area underlain by outwash sand and gravel. In both cases the soils of these areas are well drained. Spectral class five has very limited occurrence in the county, but where it does occur, it is predominantly associated with the outwash soil areas.

Spectral classes six and seven occur predominantly in the southwest and northwest parts of the county. They are very seldom found near major drainageways or in areas of rolling topography. Broad expanses of these two spectral classes occur in areas mapped as the Raub-Ragsdale association (73) and the Sidell-Parr association (89). These areas are nearly level and

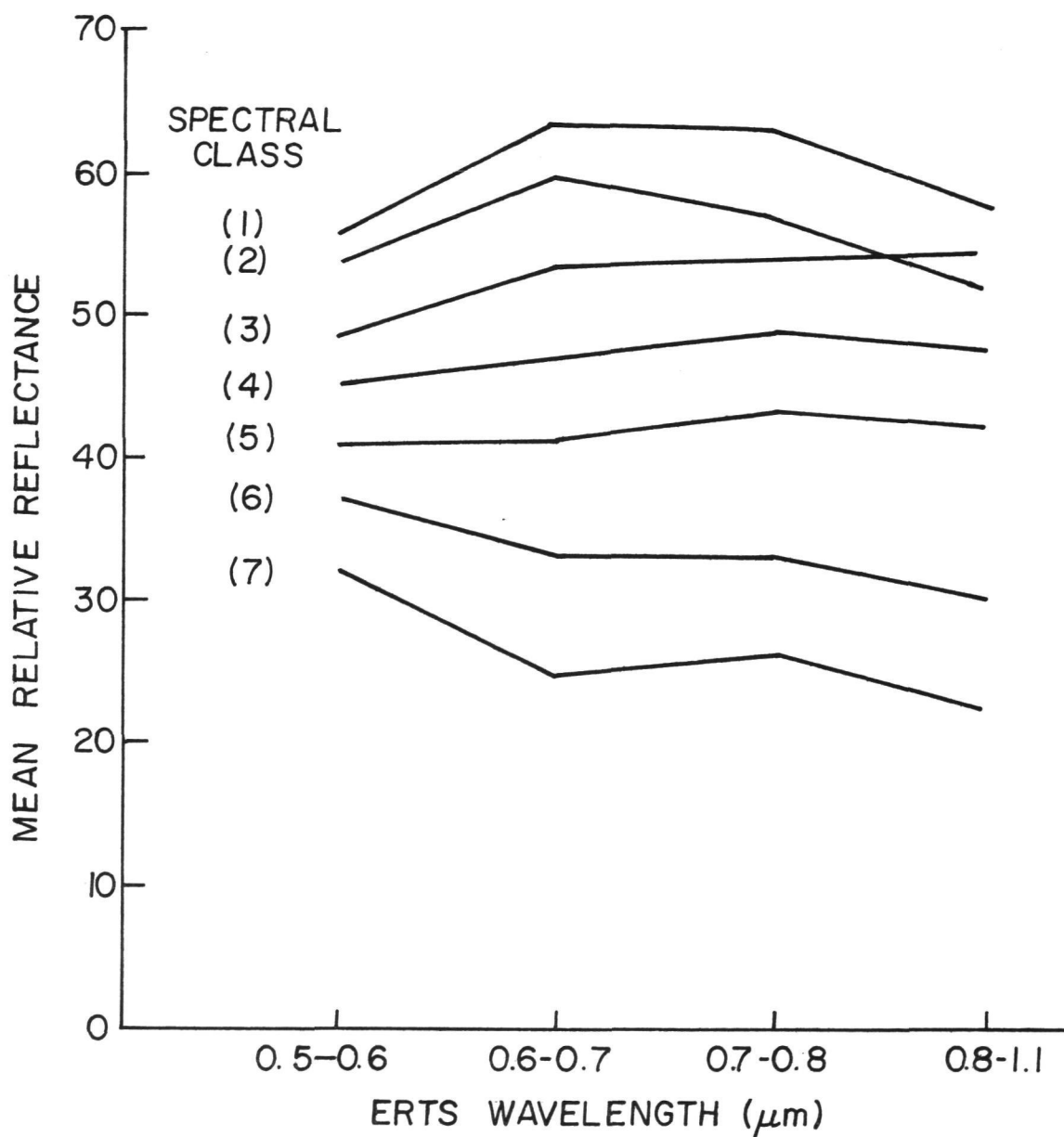


Figure 3.5 Multispectral means and variances of seven non-vegetated soil classes in Tippecanoe County, Indiana data were collected June 9, 1973.

include the somewhat poorly drained and the very poorly drained prairie soils in the case of the Raub-Ragsdale association (73). The Sidell-Parr soils are well drained prairie soils. The topography is predominantly level in the Raub-Ragsdale association (73), which is not the case for the Miami-Russell-Fincastle association (81). Spectral class six also occurs in the Elston-Wea association (16). There is a limited occurrence of spectral classes six and seven in the areas mapped as the Fincastle-Ragsdale-Brookston association (66). In the northwest part of the county it is more difficult to find broad expanses mapped purely as spectral classes six and seven.

3.33 Spectral Composition of Soil Associations

A second approach to evaluating Figure 3.4 was to examine the spectral class composition within each mapped soils association. Using this approach, it was very difficult to arrive at any 1:1 correspondence between spectral classes and soils or soil associations. This result, however, was not unexpected, since soil associations are only one of many possible interpretations of a detailed soils map. In the Fincastle-Ragsdale-Brookston association (66), for example, a broad range of surface spectral properties is included. In some areas soil association (66) consists predominantly of the somewhat poorly drained, lighter colored Fincastle soils, resulting in a predominance of spectral classes one, two, and three. Other delineations of this association consist of greater percentages of the poorly drained Ragsdale and Brookston soils. In these cases the spectral class makeup includes large amounts of spectral classes six and seven. The Fincastle-Ragsdale-Brookston association (66), had, in fact, the most variable spectral class composition of any association. In spite of this lack of 1:1 correspondence between spectral classes and soils or soil

associations, several very meaningful generalizations were made regarding spectral class makeup.

In general, the Miami-Russell-Fincastle association (81) and the Fincastle-Ragsdale-Brookston association (66) could not be distinguished from one another by their spectral class composition. Both of these consisted largely of spectral classes one, two, and three. However, the Fincastle-Ragsdale-Brookston association (66) did contain large amounts of spectral classes six and seven in addition to spectral classes one, two, and three which the Miami-Russell-Fincastle association (81) very seldom contained. That is, the Miami-Russell-Fincastle association (81) always consisted predominantly of spectral classes one, two, and three. This spectral class composition is very reasonable since the poorly drained Ragsdale and Brookston soils occur with varying frequency within the Fincastle-Ragsdale-Brookston association (66). Conversely, the Miami-Russell-Fincastle association (81) consists primarily of the sloping, well-drained soils and included some of the somewhat poorly drained soils, but the very poorly drained soils were seldom found.

A rather unique situation occurred in the spectral class makeup of one of the major outwash soil areas. There was almost no occurrence of spectral classes one and two, and a very large portion consisted of one spectral class, spectral class four, representing an intermediate reflectance. This result is very logical considering the properties of soils found in the Elston-Wea association (16). These soils are nearly level, have loamy surface horizons, and are well-drained. The soils were developed under grass-land vegetation, and are Mollisols, which therefore have darker surface colors than the soils of the Miami-Russell-Fincastle association. That is, they are darker than the well drained and somewhat poorly drained soils formed under forest

vegetation (Alfisols); however, they are not as dark as the poorly drained soils---the Brookston, Ragsdale and Chalmers.

Spectral classes one, two, and three occurred more frequently in the areas mapped as the Sidell-Parr association (89) than in the areas mapped as the Raub-Ragsdale association (73). Sidell and Parr are well drained prairie soils developed in glacial till, and generally occur on more sloping topography than the Raub and Ragsdale soils. Because soils of the Sidell-Parr association are Mollisols, the surface colors are darker than those of the predominant soils of the Miami-Russell-Fincastle and Fincastle-Ragsdale-Brookston associations (with the exception of the included poorly drained soils). It was not expected that spectral classes one and two would occur in the Sidell-Parr association (89). Conversely, it was not expected that spectral classes six and seven would predominate in these sloping areas of well drained soils. Evaluating the extent to which these kinds of expectations were observed was very difficult. In the southwest part of the county, for example, it is possible to consider the boundary drawn between the Raub-Ragsdale association (73) and the Sidell-Parr association (89) in several instances. In most of these instances, there is no observable change in the spectral class composition at this boundary. The brighter spectral classes predominate in closer proximity to the drainageways, whereas the predominance of the lower reflecting spectral classes occurs further back from the drainageways, in the more nearly level areas.

3.34 Conclusions

Both the false color imagery and the computer classification produced useful information relating to the soils of Tippecanoe County. Agreement between ERTS imagery and the conventional generalized soil map was noted in several instances. In cases

where lack of agreement was noted, ground observations and examination of aerial photography indicated discrepancies in the generalized soil map. The ERTS sensors were definitely sensitive enough to detect meaningful differences in surface reflectance characteristics of the various soils. Relating these surface spectral characteristics to the soil properties which are of interest to soil surveyors increased the complexities of the investigation.

3.4 Finney County, Kansas

In viewing the ERTS red-band imagery of Finney County, Kansas, several soil associations are easily delineated. In Figure 3.6 soil association 3 (Richfield-Ulysses-Mansic), which consists of loamy soils of the Pawnee River drainage basin, appears as the darker area in the northeast part of the county; drainage patterns are visible throughout. The drainage patterns are the darkest because they are lined with various vegetation which is a healthy green and would appear dark in this red-band imagery. The other dark areas represent bare soils as do the lighter areas. After careful examination and comparison of the imagery with a topographic map; the dark areas are noted as being the steeper sloping soils and the lighter colors are on higher ground and more nearly level. Richfield, then would be typical of the lighter color in this association being a gently sloping silt loam. These soils are well suited for wheat and grain sorghum, and they present no major erosion problems when managed properly. Ulysses, then, is represented as the darker patterns having up to a 5% slope. Major management problems with this soil are conserving moisture and controlling wind and water erosion. The very steepest soil is the Mansic with up to 15% slope. These are not suitable for cultivation because they are too steep and susceptible to erosion. They are kept in grass

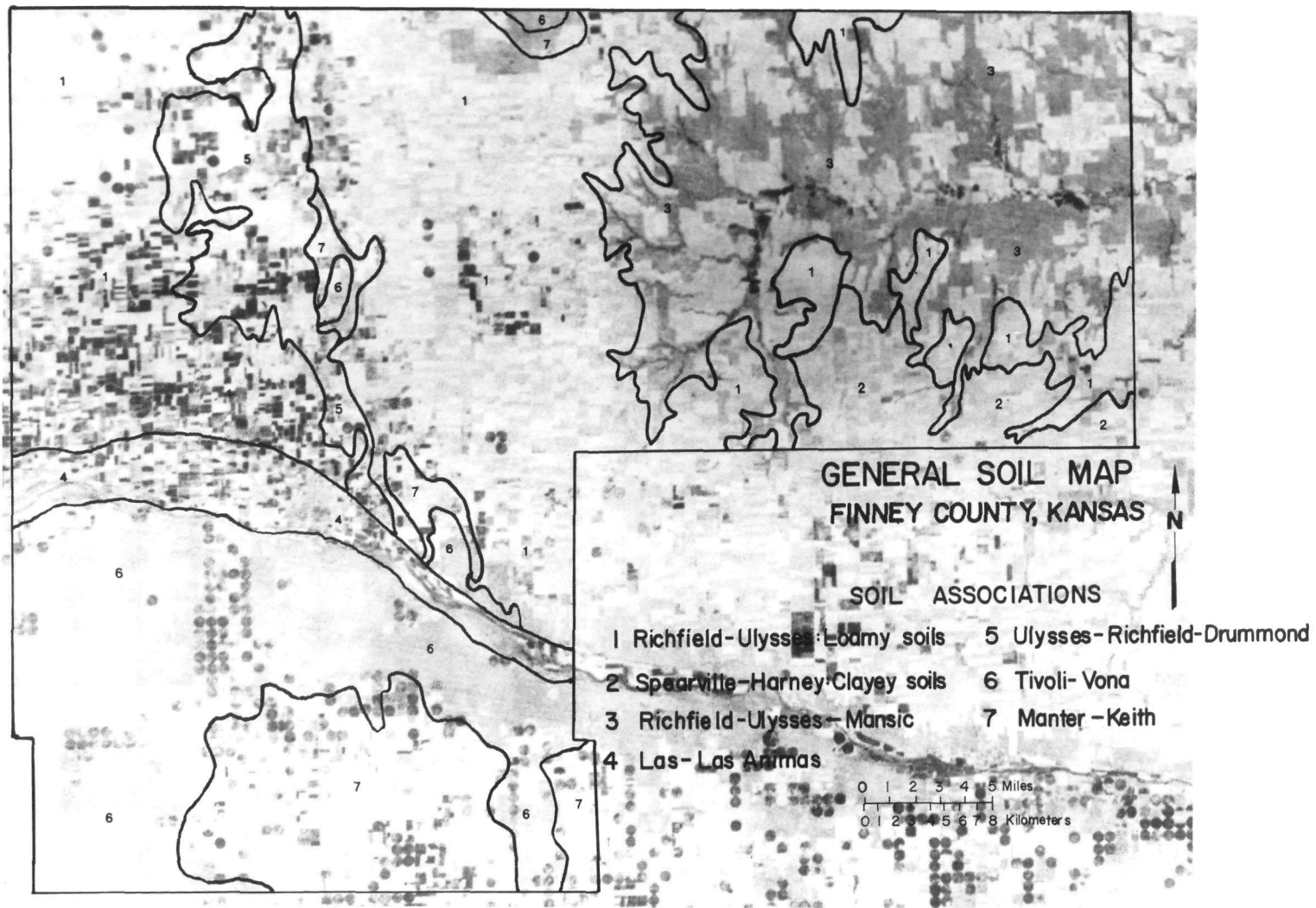


Figure 3.6 Black and white reproduction of ERTS data from channel two taken July 6, 1973 over Finney County, Kansas with an overlay of the county soil association map.

for grazing. So as previously mentioned pertaining the drainage patterns, these steeper sloping soils are supporting healthy green vegetation. Humbarger, the flood plain soil in this area, would also be covered with vegetation and thus also be darker in the image.

Next under consideration will be the mapping of the association 5 (Ulysses, saline-Richfield, saline-Drummond), the soils of the Scott-Finney depression. The soils of this association are defined as being of a depression and are mapped accordingly. The boundary of this soil association in general follows the topographic contour line at an elevation of 2,875 feet, which is lower than the surrounding area. However, there is just a very slight slope (approximately .1%) to the west of this boundary and a much more abrupt slope noticed to the east. Thus, it is not possible to note any difference in features to the west from the depression, as opposed to the steeper boundary of the high plains to the east.

These topographic features are implied by the imagery. It can be noted that farmers have cultivated the area north of the Arkansas River from the Western County boundary through association 1, through the Scott-Finney depression, and halt rather abruptly at the eastern boundary of the depression. This is in view of the fact that the slope to the west is slight so the problems of wind and water erosion are minimal; however, to the east the soils are steeper and more susceptible to erosion.

The next association that will be discussed is number 7 (Manter-Keith), the majority of which is located in the southern portion of the county. These are the sandy and loamy soils between the sand hills and the table lands. In the image this area appears generally lighter than the areas surrounding. When compared to a topographic map this area is noted as being generally level and the western boundary contours perfectly with the boundary of the sand hills of association 6.

Association 6 (Tivoli-Vona), the soils of the sand hills appear as the grayish toned area around association 7. Circular areas are noted in this area. These are irrigated areas with a diameter of one half mile. Irrigation is the only way to get crops to grow in this sandy area.

3.41 Conclusions

From this study of Finney County, Kansas, a strong relationship was found between ERTS imagery (single band only) and soil associations as mapped by conventional means. It is anticipated that a greater amount of soils information could be extracted from the ERTS data using multiple bands and computer processing.

3.5 Summary and Conclusions

Strong relationships between ERTS imagery and conventionally mapped generalized soil boundaries were found for all three test counties. Results indicate that computer analysis of MSS imagery provided better discrimination among soils than single band imagery or false color enhancement using multiple bands. A major limitation of the computer analysis was selection of training samples which were representative of the soils across the county. The approach used in the Tiptecanoe County study was to enhance the strong spectral differences between soils over a large area (501 square miles). Computer analysis of MSS data was more flexible than the photographic approach in several respects: (1) It facilitated analysis over smaller areas in more detail. Previous studies of smaller areas indicated that it was possible to separate spectrally unique classes with greater precision than was done in the study reported here. (2) It was possible to select a data set from a small area (such as a county) rather than using an entire frame of data. This subset of spectral data could then be histogrammed, resulting in greater contrast between

features of interest in the study area than was obtainable from the original data over the entire frame. This improvement in discrimination of spectral classes in the scene has been found to increase separability, however, such comparisons are not presented in this report. (3) The use of three spectral bands and simulation of color infrared photography have been shown to enhance differences among soils to a greater extent than single band photographic techniques. Comparisons have been previously reported by other researchers. Use of color IR simulation techniques are subject to many of the same limitations which are inherent with conventional color and color IR aerial photography. These shortcomings of aerial photography in remote sensing are well known. Among other considerations, only linear combinations of wavelengths are possible using the false color enhancement of ERTS MSS data and the interpretations which can be made are still largely subject to tonal differences rather than actual measured differences in multispectral reflectance.

Many more advantages of digital computer processing of ERTS multispectral data for soil survey purposes will be demonstrated in the near future. A capability which should not be overlooked is that the acreages of each spectral class can be estimated readily from the results of the computer analysis. A similar quantitative tabulation cannot be easily obtained from photography. To the extent that a spectral class can be related to a soil type or condition, it will be possible to estimate acreages of soils within a given geographic area. A further capability which was not employed in this study was utilizing ground information about the soil types to "train" the computer to predict other soil characteristics using only spectral characteristics. This is the so-called "supervised" training approach. The analysis performed here did not utilize

ground information in the computer analysis phase of the study, except to ascertain that only bare soil spectra were included in the spectral data set analyzed. It has been found from many previous studies of this type that if ground information can be adequately determined, a greater accuracy of classification of ground cover types is obtained from using the supervised (training) approach, in comparison with the unsupervised (clustering) approach which was used in this study.

Studies are presently underway to evaluate the usefulness of computer processed ERTS imagery in the ongoing Cooperative Soil Survey Program. These studies involve soil scientists who are presently engaged in field mapping, and the use of computer processed ERTS imagery for gaining additional information during the survey. Preliminary investigations indicate that ERTS imagery will be useful early in the planning stages of the survey. Studies are being undertaken to test these procedures and several types of ERTS computer generated products in a variety of vegetative, climatic, and soil conditions across the central United States. The usefulness of these techniques in low and medium order reconnaissance soil surveys seems appraent at this time. It is believed that improved field mapping techniques will result from the use of computer-processed ERTS multispectral imagery in operational soil mapping. This should greatly impact the techniques which can be used in accelerating the present soil survey program.

Preliminary studies which were conducted in the late 1960's as remote sensing was becoming more involved in soil mapping pointed out that soil moisture, soil surface roughness, and other surface conditions not directly related to the mapping of soils had some effect on soil spectral characteristics. In this study no information was obtained as to the surface condition at the time the spectral data were gathered, other than to assure that the surfaces were nonvegetated. The results obtained

from this study are particularly encouraging when it is considered that while soil surface conditions were confounded with soil properties of interest in mapping, it was still possible to separate soils into meaningful classes over a large area.

4.0 URBAN LAND USE ANALYSIS

4.1 Introduction

The urban land use analysis project was designed to test the ability of ERTS-1 data and multispectral pattern recognition techniques to recognize and map significant features in the urban scene. Land use planning officials would benefit greatly from the acquisition of land use information and monitoring of changes in land use in their areas of jurisdiction from the ERTS-1 satellite data. Presently, large metropolitan areas complete land use studies only once every several years. Such inventories are very costly, in terms of man-hours invested, and are often critically obsolete by the time they are completed. Because the satellite passes over any given area once every eighteen days, changes in land use may be monitored on a timely basis.

Since the launch of the ERTS satellite, researchers have analyzed the ERTS-1 multispectral scanner data in two major fashions. One approach has been pictorial. The analyst observes differences in the spectral reflectance of land cover types by noting differences in gray levels in the imagery. He either studies the four band images separately or uses color composites (color enhanced images produced by a combination of three of the four ERTS bands). The second approach utilizes computer analysis, and is referred to as the digital, or, numerical approach. The four bands of data are digitized, spatially registered, and stored on magnetic tape in computer compatible format. The digital approach was used in the analysis reported here.

The underlying assumption of the digital approach is that a certain land cover type is spectrally separable, i.e., it has a unique range of spectral responses in one or more but not

necessarily all, of the spectral bands collected. If the assumption is true, then pattern recognition computer programs can be used to correctly identify the particular land cover type. Research completed with multispectral scanner data collected from aircraft has shown that important earth surface features, such as roads, rooftops, grass, trees, water, and bare soil can be accurately identified.

Whereas detailed land cover studies can be made from aircraft altitudes, gross patterns of urban land use are detected in ERTS analyses*. The IFOV, i.e., the instantaneous field of view, of the satellite's multispectral scanner is approximately 1.1 acres (0.45 hectares). Consequently, a single remote sensing unit (RSU) from ERTS collected over a residential area, for example, includes rooftops, streets, grass, trees, and shrubs. The scanning device integrates the various spectral responses detected from these land cover types, resulting in a single response for each RSU and for each spectral band.

* Ellefsen, R., Swain, P. H., and Wray, J. R., 1973. Urban land use mapping by machine processing of ERTS-1 multispectral data: A San Francisco Bay area example. Proceedings of Conference on Machine Processing of Remotely Sensed Data. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, October 16-18, 1973. P. 2A-7 to 2A-22.

Todd, W. J., Mausel, P. W., and Wenner, K. A. Preparation of urban land use inventories by machine-processing of ERTS MSS data. Proceedings of Symposium on Significant Results Obtained from the Earth Resources Technology Satellite-1. Goddard Space Flight Center, Greenbelt, Maryland, March 5-7, 1973. Vol. I, Sec. B, p. 1031-1039.

Todd, W. J., and Baumgardner, M. F., 1973. Land use classification of Marion County, Indiana by spectral analysis of digitized satellite data. Proceedings of Conference on Machine Processing of Remotely Sensed Data. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, October 16-18, 1973. P. 2A-23 to 2A-32.

4.11 Scope of Report

Urban land use analyses are reported herein for four large metropolitan areas: Milwaukee and Chicago (Frame ID 101716093), Indianapolis (Frame ID 106915585) and Gary, Indiana (Frame ID 107016050). Initially, the Milwaukee subframe was analyzed (Section 4.2). Statistics from the Milwaukee area were then used to classify the Chicago subframe, to test the reliability of training sets (Section 4.3). Analysis of the Indianapolis subframe was performed next (Section 4.4) and lastly the Gary area was analyzed (Section 4.5). Important lessons were learned from the Indianapolis study. Thusly, it was decided to refine the Milwaukee - Chicago classifications (Section 4.5). The Gary analysis benefited greatly from the earlier three studies.

Use of geometrically corrected ERTS data is reported in Section 4.6 and use of histograms in conjunction with the digital display in Section 4.7. Analysis of the Gary, Indiana data is discussed in Section 4.8. Conclusions are made in Section 4.9.

4.2 Milwaukee County Subframe Analysis

4.21 Data Processing

Initially, the four bands of Milwaukee data were examined on the LARS digital imaging display. Several important functions were performed at that time. One, the county boundaries were determined, along with the line/column coordinates of a number of landmarks (e.g., interstate highways, airports, lakes, rivers) to facilitate the interpretation of line printer maps in subsequent steps in the analysis. Two, areas were chosen for the histogramming processor to obtain histogram decks for maximum contrast viewing of the study area in future examinations

of the study area on the digital display. Three, the quality of the data was examined. It was immediately noticed that "six-line noise" was evident in Band 4, especially in water. Four, a small area in the central part of the county was chosen for clustering.

Fourteen spectral classes were requested of the clustering processor. Statistics were then calculated for the clusters delimited, and the entire study area was classified. The resulting non-supervised classification of Milwaukee County was not satisfactorily representative of the land uses present. Nevertheless, the cluster map did provide a base map from which rectangular "Training samples" for the desired land use classes could be chosen manually. One class chosen by the clustering algorithm, representing grassy, agricultural areas, was retained. Combining that class with the other classes chosen manually, a new set of statistics was calculated and the study area was classified again.

4.22 Areal Distribution of Classes

4.221 Explanation of Classification Image

The classification results were photographed from the digital imaging display (Figure 4.1). Graylevels used for the display of the spectral classes are as follows:

| | | |
|-----------------|---------------|---------------------------|
| Road - Downtown | - white | Water 1 - dark gray |
| Industry | - medium gray | Water 2 - very dark gray |
| Inner City | - black | Water 3 - black |
| Suburban | - white | Water 4 - black |
| Wooded suburban | - light gray | Water 5 - very light gray |
| Grassy 1 | - dark gray | Cloud - white |
| Grassy 2 | - dark gray | Shadow - black |



Figure 4.1 Computer-implemented land use classification of Milwaukee County, Wisconsin.

Certain classes have been assigned the same gray levels, but consideration of their areal distribution allows visual separation. Water 3, Water 4, Inner City, and Shadow are all displayed as black. Water 3 is located only in Lake Michigan, and is the most Eastward of the succession of water classes into Lake Michigan. Water 4 is found principally in two large lakes, Lake Muskego and Little Muskego Lake (Southwestern part of study area) and in Milwaukee Harbor (in Lake Michigan) in the central portion of the study area. The class Inner City is found in the central part of the study area, surrounded by suburban areas (white) and Water 1 (dark gray). Cloud shadows have very limited distribution, and are found in the outer areas of the study area. They are small, and each is associated with a cumulus cloud (white) located a small distance to the Southeast.

Similarly, Suburban, Road - Downtown, and Cloud have all been displayed as white. The distribution of Suburban and Cloud was discussed above. Road - Downtown is located in the central part of the image, surrounded by Inner City (black).

Finally, Water 1 and Grassy areas are both displayed as dark gray. Water 1 is the water class in Lake Michigan located along the shore (except when Water 5 occupies a very narrow band of water along shore), while grassy areas are located in the outer parts of the county, North, West, and South of suburban areas.

4.222 Location and Characteristics of Spectral (land use) Classes

The white area in the central part of the study area (classified as Road - Downtown) contains Milwaukee's Central Business District. This class is almost totally a mixture of

rooftops and concrete, and thus has a very high reflectance in the visible bands. Data points of this class also occur along interstate highways and sandy beaches.

The first major ring of land use outward from the Central Business District, displayed as black, is termed "inner city". This class extends from approximately Burleigh Street on the North to Cleveland Street on the South, and from 60th Street on the West to Lake Michigan on the East. A great many of the homes in this area are the bungalow or "two-flat" (multiple-story) type of structure, most of which were built prior to 1940.

Typically, the houses were built very close together, and are inhabited by two or more families. Mature vegetation (large trees) and closely spaced rooftops are the primary constituents of this spectral class.

The class "industry" (shown as medium gray) was identified only where the larger areas of heavy industry predominate. Two large industrial areas were identified, which together form an L-shaped region, located just South of the Central Business District. One is in the Menomonee River Valley and the other in the Kinnickinnick River Valley. Five smaller industrial areas identified include the Capital Drive - 35th Street area, Capital Drive - Richards Street area, 70th Street - Greenfield area (in West Allis), Southern West Milwaukee, and Packard Avenue (in Cudahy). Industrial areas are characterized by a large proportion of rooftops with surrounding roads. Data points of the class "inner city" are frequently interspersed among the industrial areas.

The ring North, West, and South of the "inner city" is an area of complex land uses, including suburban, recreational, and institutional land uses. The three primary cover types in this area are "suburban", "wooded suburb", and "Grassy".

"Suburban" areas (shown as white) are dominated by single-family dwelling units built on moderately sized lots. Most of the structures were built after World War II. Roads and lawns (grass) are the two land cover types which most nearly characterize this spectral class. The principal areas are the outer areas of the City of Milwaukee, Northern Wauwatosa, West Allis, Greenfield, Greendale, Hales Corners, Cudahay, South Milwaukee, St. Francis, and Brown Deer.

The areas of "wooded suburb" (displayed as light gray) include Southern Wauwatosa, Fox Point, Whitefish Bay, and Shorewood. This class consists of old, single-family dwellings built on large wooded or grassy lots. These areas are the older, upper-income areas of Milwaukee County, located only four or five miles of the Central Business District (CBD). They are not to be confused with newer, upper income areas found some ten or more miles from the CBD, and constructed after World War II.

"Grassy" areas in this ring manifest themselves as parks, golf courses, and cemeteries.

The final ring of land use, outward from "suburban" areas, were classified as "grassy". Agriculture is the dominant land use, although a number of "grassy" areas are plots of idle land, probably owned for speculative reasons by land developers. Trouble was encountered in this ring in the classification of a number of very recently developed subdivisions (post-1960), particularly in the Brookfield, Elm Grove, and New Berlin municipalities (all three are in Waukesha County, West of Milwaukee County). Such residential areas were mis-classified as "grassy".

Five spectral classes of water were identified within the study area. Four of the classes, "Water 1", "Water 2", "Water 3", and "Water 5" are located almost exclusively in Lake Michigan.



Figure 4.2 Computer-implemented land use classification of Chicago area. Black line is boundary of Cook County.

There is a regular succession of water classes Eastward into the lake. Though this suggests that the water classes are indicative of depth, examination of two U. S. Geological Survey topographic quadrangles indicates little association between the water classes and depth. The fifth class of water, "Water 4", occurs in small water bodies (such as Lake Muskego and Little Muskego Lake) and in the Milwaukee River. Not unexpectedly, this class also appears in Milwaukee Harbor and along Lake Michigan's coast to the South. Evidently, water from Milwaukee Harbor (into which flows the Milwaukee River) slowly mixes with the lake water as it is carried South along the coast. Factors explaining the five spectral classes of water may be variations in color and turbidity.

4.3 Chicago Subframe

4.31 Data Processing

Analysis of the Chicago subframe was undertaken to determine if the spectrally separable classes (statistics deck from LARSYS *STATISTICS computer program) which were used successfully in one metropolitan area could produce satisfactory results in another metropolitan area. The Chicago spectral data (from the same ERTS frame) were classified using the Milwaukee statistics, and the results photographed from the digital imaging display (Figure 4.2). The same gray levels for classes are used for the Chicago classification as were used for the Milwaukee analysis. Classification results for Chicago similar to those of Milwaukee were achieved.

4.32 Areal Distribution of Classes

The first major ring of land use outward from the Central Business District was classified as "inner city". This area included most of the City of Chicago, along with Cicero, Berwyn, and Blue Island. Larger parks, cemeteries, and large industrial

areas were identified within this ring.

The next ring outward included large areas of the classes "suburb" and "wooded suburb". Many small, scattered areas of the class "suburb" appeared, but the larger ones were found in the municipalities of Oak Lawn, Hodgkins, Norridge, Harwood Heights, and Morton Grove. Six major regions of the class "wooded suburb" were:

1. Glencoe, Winnetka, Kenilworth, Wilmette, Evanston
2. Park Ridge
3. Elmwood Park, Oak Park, River Forest
4. Elmhurst, Villa Park, Lombard, Glen Ellyn, Wheaton
5. Riverside, Brookfield, LaGrange, Western Springs, Hinsdale, Golf View Hills, Westmont, Downers Grove
6. An area bounded by the Dan Ryan Woods on the North, 119th Street on the South, Beverly Street and Vincennes on the East, and Western Avenue of the West.

The classes "cloud" and "shadow" were identified in the Northwestern part of the Chicago area, North of Chicago O'Hare International Airport.

4.4 Marion County Subframe

4.41 Introduction

Upon completion of a reasonably successful analysis of the Milwaukee subframe, a logical sequence of investigation was the attempt at replication of results in another metropolitan area. Marion County (Indianapolis) was the area selected for the second urban land use analysis.

4.411 Similarities Between the Milwaukee and Indianapolis Subframes

Significant differences in results of analyses of the Milwaukee and Indianapolis subframes were not expected. Both metropolitan areas are located in the "Midwestern" region of the United States, and both cities, consequently, have experienced the same patterns of urban development. Prior to World War II, urban growth was largely contained, i.e., residential expansion was contiguous with the already existing built-up area. Commercial and industrial activity was restricted largely to the center of the urbanized area. Residential land use surrounded this commercial/industrial core, and consisted of closely spaced structures with few "open" (unused) lots. After World War II, urban areas experienced unprecedented construction of subdivisions, or, suburban areas. Lack of urban planning, increasing use of the automobile, and land speculation all contributed to the "leap-frogging" pattern of development. Subdivisions were located wherever the developer could buy inexpensive land; farmland and idle land, consequently, often lay between suburban developments. Industrial developers, moreover, followed the subdivisions.

Climatically, the two urban areas are also very similar. Both have humid, continental climates. Milwaukee has colder Winters and cooler Summers, however, than does Indianapolis. Milwaukee is classified (according to Koeppen) as Dfb; Indianapolis as Dfa. Topographically, both urban areas are relatively flat and have limited relief.

4.412 Differences Between the Milwaukee and Indianapolis Subframes

The most significant difference between the two data sets is the date of data collection; the Milwaukee ERTS pass was 9 August, while Indianapolis was 30 September. The Milwaukee data were collected when virtually all areas of vegetative cover were very green, while the Indianapolis data were collected in early

Fall, when a significant proportion of agricultural fields were either bare soil or consisted of browning vegetation.

Other important differences are varying latitudes (Milwaukee is 43.05 degrees of latitude North; Indianapolis 39.43 degrees North) and differences in sun angle between the two dates.

4.42 Data Processing

The technique of analysis varied little from the Milwaukee study to the Indianapolis study. Initially, the data were examined on the digital display. The county boundary was determined, areas were chosen for histogramming, and areas were chosen for clustering. Basic spectral groupings in the data were found by using the clustering algorithm, and the resulting cluster map of the county was used to pick manually a set of training fields for the desired land use classes. Statistics were calculated for the classes, and the study area was classified.

4.43 Classification Results

4.431 Explanation of the Classification Image

The classification results (as photographed from the digital display) are shown in Figure 4.3. Gray levels used for display of the spectral classes are as follows:

| | |
|-------------------------------|---------------|
| Commerce/Industry | - medium gray |
| Inner City | - black |
| Suburban | - white |
| Wooded | - light gray |
| Grassy (open or agricultural) | - dark gray |
| Water | - black |
| Cloud | - white |
| Cloud Shadow | - black |

Several pairs or trios of classes have been given the same gray level, but consideration of their areal distribution permits visual separation. "Suburban" and "clouds" are both white, but the clouds are all small, of the cumulus variety, and have an associated shadow located approximately one kilometer to the Northwest. "Inner City", "Water", and "Cloud Shadow" are all displayed as black, but visible separation is also possible. "Cloud Shadows" are associated with the cumulus clouds. "Water" is largely limited to two large reservoirs, Eagle Creek (West-central portion of image) and Geist (Northeast corner), and to several large ponds. "Inner City" is located in the center of the county, surrounding the Central Business District (classified as commercial/industrial).

4.432 Areal Distribution and Characteristics of Spectral Classes

"Suburban" (displayed as white) is a class consisting of residential areas developed primarily after World War II, similar to the suburban class used in the Milwaukee study. Housing density is relatively low and family incomes moderate. Three large areas were classified as "suburban".

APPROXIMATE BOUNDARIES

| | <u>NORTH</u> | <u>SOUTH</u> | <u>EAST</u> | <u>WEST</u> |
|----------|--------------|-----------------|---------------|----------------|
| 1. WEST | 46th Street | 10th Street | Tibbs Ave. | I-465 |
| 2. EAST | 62nd Street | Washington St. | Church Rd. | Arlington Ave. |
| 3. SOUTH | Edgewood Rd. | County Line Rd. | McFarland Rd. | Bluff Rd. |

Streets and lawns (grass) are the two primary cover types responsible for the spectrally separable nature of this class. Not unusually, therefore, interstate highways, boulevards, and airport runways were classified as "suburban".

The class "Commerce/Industry" (displayed as medium gray) is characterized by the occurrence of rooftops, streets, and parking

lots. In the Milwaukee study, the attempt had been made to separate commerce and industry (the "Road - Downtown" and "Industry" spectral classes), but combining the two in Indianapolis resulted in much better classification accuracy. The largest area classified as commerce/industry is the Central Business District of Indianapolis (central portion of image) and adjacent industrial areas. This area extends from approximately 20th street on the North to Morris Street on the South, and from West Street on the West to College Avenue on the East. Other, smaller areas in the outer part of the city were also classified as such; they include larger industrial establishments and shopping centers. All areas in this spectral class are typified by a lack of green vegetation.

The class "inner city" (shown as black) in Indianapolis occurs as a ring of land use surrounding the Central Business District. The ring is bounded by 56th Street on the North, Troy Avenue on the South, Tibbs Avenue on the West, and Arlington Avenue on the East. At least 75 per cent of the structures in this area were built prior to World War II, not dissimilar to the age of housing in the "inner city" area of Milwaukee. Mature tree cover is a primary influence in the spectral responses from these areas, as are the closely spaced rooftops.

"Grassy" (open or agricultural) areas, shown as dark gray, are found in the outer part of the county. This class includes cropland, pasture, and idle land in rural areas, as well as grassy features in urban areas, such as parks, golf courses, and cemeteries. Areas classified as "trees" (a class not obtained by analysis of the Milwaukee data), displayed as light gray, are closely associated with the drainage pattern of the county. The most extensive stands of trees are located around Geist and Eagle Creek Reservoirs. The distribution of "Water", "Cloud", and "Cloud Shadow" was discussed above.

4.44 Accuracy of Classification

An attempt was made to assess the accuracy of the Indianapolis classification by a sampling method. Several rectangular areas, termed test fields, were located for each of the spectral classes and the class accuracy determined (Table 4.1). Four of the eight classes -- "Commerce/Industry", "Suburban", "Woodland", and "Water", -- were identified with over 90 per-cent accuracy. "Cloud", "Cloud Shadow", and "Inner City", had correct recognitions in the 80 to 90 per-cent range. "Grassy" (open or agricultural) areas were the most poorly identified -- only 64.5 per-cent correct recognition. Overall classification accuracy (the mean of the eight values) was 87.1 per-cent. Elimination of error due to weather conditions at the time of data collection (cloud and cloud shadow classes) raises the accuracy slightly to 87.5 per-cent.

The classification accuracy was achieved utilizing only spectral information. No attention was given to areal information in the data, i.e., theoretical considerations of urban geography, growth, and planning. Areal information could be introduced into the scheme of classifying urban land use. For purposes of simplification, the spectral classes may be divided into two general categories -- urban and rural. The urban category would include the classes "Commerce/Industry", "Inner City", and "Suburban"; rural would include "Wooded" areas, "Grassy" (open or agricultural), and "Water". Boundaries could be stored in a computer, delineating the urban-rural boundary in Marion County. Data points within an urban area, for example, could only be classified into one of three classes, "Commerce/Industry", "Inner City", or "Suburban". Applying this theory to the test results in Table 4.1 gives the values in the extreme right column. Accuracies for each class are greater than 90 per-cent, and the overall classification accuracy has been increased to 96.4 per-cent.

Table 4.1. Accuracy of Classification

| Spectral Class | Percentage of data points classified into: | | | | | | | | % with ⁵ areal |
|-------------------------------|--|------------------|------------------|-------------|-------------------|-------------|-------------|-------------|------------------------------|
| | C/I ¹ | OHg ² | NHg ³ | Wood | Grsy ⁴ | Cld | CdSh | Watr | |
| 1. Commerce/Industry | <u>96.0</u> | 1.1 | 1.3 | ---- | 1.6 | ---- | ---- | ---- | 97.6 |
| 2. Older housing ² | ---- | <u>81.0</u> | 8.2 | 1.0 | 9.7 | ---- | ---- | ---- | 91.7 |
| 3. Newer housing ³ | 0.2 | 2.0 | <u>91.2</u> | ---- | 6.0 | ---- | ---- | ---- | 97.2 |
| 4. Wooded | ---- | ---- | 0.3 | <u>99.4</u> | 0.3 | ---- | ---- | ---- | 99.7 |
| 5. Grassy ⁴ | 7.9 | 25.4 | 2.5 | 6.6 | <u>64.5</u> | ---- | 0.1 | ---- | 93.2 |
| 6. Cloud | ---- | ---- | 14.4 | ---- | ---- | <u>85.6</u> | ---- | ---- | |
| 7. Cloud shadow | 11.5 | 0.7 | ---- | ---- | ---- | ---- | <u>86.3</u> | 1.4 | |
| 8. Water | 3.3 | 2.9 | ---- | 0.9 | ---- | ---- | ---- | <u>92.9</u> | 99.1 |
| Ponds | 2.6 | ---- | ---- | 0.3 | ---- | ---- | ---- | 97.3 | 99.8 |
| Streams | 16.7 | 47.9 | ---- | 10.4 | ---- | ---- | ---- | 25.0 | 89.6 |

Overall classification accuracy = 87.1%

Accuracy minus weather conditions = 87.5%

(minus cloud and shadow)

Accuracy with areal information = 96.4%

(minus weather conditions)

¹Commerce/Industry

²Multi-family (older) residential

³Single-family (newer) residential

⁴Grassy (open, agricultural) areas

⁵Percentage with areal information (urban-rural differentiation)

Most of the error in classification was attributed to the confusion between "Grassy" (open or agricultural) areas and "Inner City" (Table 4.1). Other problems of classification arose in two types of residential areas, neither of which could be separated as single spectral classes. One of these types may be referred to as a transitional residential area. It is located between areas classified as "Inner City", with 75 per-cent or more of its structures having been built prior to World War II, and "Suburban", with 25 per-cent or less of its structures built prior to the second world war. The transitional residential areas have housing of mixed age, 25 to 75 per-cent of its structures having been built prior to World War II. The second type of residential area is found in the North-central part of the county, from County Line Road South to 56th Street and from Northwestern Avenue East to Interstate 465. Within this area are scattered residential developments, built after World War II, and consisting of upper income families. Such areas are termed vegetative residential.

Classification results were not satisfactory for the above four types of land use -- grassy, inner city, transitional residential, and vegetative residential -- using the Gaussian maximum likelihood classifier, but evaluation of certain parameters (Table 4.2) did allow separation of three of the four land uses by sample. The means and standard deviations in the visible bands of the spectrum presented no evidence of separability between the land uses. In the infrared bands, however, certain of the land uses are separable. Vegetative residential is readily separable from the other two residential land uses, because of the higher reflectance of the former. Inner city and transitional residential are not separable by application of these parameters.

Although sample means do not indicate separability of grassy samples from the other land use samples, consideration of the

Table 4.2a Quantitative Information for Samples from
Four Selected Land Uses for All Four ERTS
Bands¹ (means and Standard Deviations)

| Land ² Use | | Means and Standard Deviations | | | | | | | |
|--------------------------|---|-------------------------------|-------------------------|-------------|------------|-------------|------------|-------------|------------|
| | | 4 \bar{X} ³ | 4 σ ⁴ | 5 \bar{X} | 5 σ | 6 \bar{X} | 6 σ | 7 \bar{X} | 7 σ |
| Tr. Res. | 1 | 24.50 | 2.36 | 17.09 | 2.62 | 32.38 | 2.25 | 18.94 | 1.14 |
| | 2 | 24.24 | 2.60 | 16.44 | 2.89 | 35.14 | 3.56 | 20.97 | 2.26 |
| | 3 | 24.33 | 2.01 | 16.73 | 2.17 | 32.81 | 2.92 | 19.06 | 1.78 |
| OHg | 1 | 27.67 | 1.79 | 21.50 | 1.82 | 32.46 | 1.96 | 17.94 | 1.06 |
| | 2 | 26.57 | 1.75 | 20.29 | 2.22 | 29.33 | 2.03 | 15.76 | 1.48 |
| | 3 | 25.85 | 1.23 | 19.75 | 1.29 | 31.80 | 1.79 | 16.95 | 0.83 |
| | 4 | 27.11 | 1.81 | 20.83 | 1.92 | 31.06 | 1.76 | 16.89 | 0.90 |
| | 5 | 26.27 | 2.14 | 19.58 | 2.60 | 30.09 | 1.96 | 16.39 | 0.83 |
| | 6 | 24.74 | 1.81 | 18.26 | 2.05 | 29.15 | 1.63 | 14.93 | 1.14 |
| Veg. Hs. | 1 | 27.38 | 1.95 | 20.12 | 2.68 | 40.38 | 2.17 | 23.31 | 1.52 |
| | 2 | 25.17 | 1.62 | 17.73 | 2.07 | 41.10 | 2.59 | 24.33 | 1.95 |
| | 3 | 27.29 | 1.16 | 20.17 | 1.63 | 42.17 | 2.58 | 25.04 | 1.68 |
| | 4 | 25.72 | 1.18 | 18.39 | 1.42 | 41.89 | 2.61 | 25.11 | 1.32 |
| | 5 | 27.75 | 1.96 | 21.75 | 1.86 | 43.58 | 1.51 | 24.75 | 1.06 |
| | 6 | 27.25 | 1.14 | 20.00 | 1.71 | 42.08 | 2.35 | 25.17 | 1.70 |
| Grassy | 1 | 24.06 | 1.56 | 19.14 | 2.11 | 30.60 | 6.85 | 17.27 | 4.85 |
| | 2 | 24.18 | 1.63 | 18.45 | 2.15 | 31.87 | 6.89 | 18.27 | 4.86 |
| | 3 | 24.82 | 2.44 | 19.36 | 4.41 | 34.58 | 6.53 | 20.10 | 5.16 |
| | 4 | 22.78 | 1.72 | 16.73 | 2.12 | 31.64 | 6.13 | 18.43 | 4.48 |
| | 5 | 23.87 | 1.37 | 18.64 | 2.36 | 32.76 | 8.00 | 18.88 | 6.05 |
| | 6 | 23.96 | 2.04 | 17.75 | 2.63 | 33.63 | 7.56 | 19.51 | 5.11 |

¹Band 4, 0.5-0.6 μ m; Band 5, 0.6-0.7 μ m; Band 6, 0.7-0.8 μ m; Band 7, 0.8-1.1 μ m.

²Tr. Res. = transitional residential; OHg = multi-family (older) residential;
Veg. Hs. = vegetative residential; Grassy = grassy (open, agricultural)
areas.

³Relative mean spectral response for Band 4.

⁴Standard deviation for Band 4.

⁵Correlation between Band 4 and Band 5.

Table 4.2b Quantitative Information for Samples from
Four Selected Land Uses for All Four ERTS
Bands¹ (correlation coefficients)

| Land ² Use | | Correlation Coefficients | | | | | |
|--------------------------|---|------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | | r ₄₅ ⁵ | r ₄₆ | r ₄₇ | r ₅₆ | r ₅₇ | r ₆₇ |
| Tr. Res. | 1 | +.88 | +.49 | +.00 | +.42 | -.06 | +.59 |
| | 2 | +.87 | +.43 | +.03 | +.31 | -.14 | +.79 |
| | 3 | +.77 | +.42 | +.18 | +.39 | +.07 | +.82 |
| Oilg | 1 | +.73 | +.30 | +.08 | +.33 | +.16 | +.61 |
| | 2 | +.79 | +.63 | +.60 | +.42 | +.34 | +.83 |
| | 3 | +.52 | +.28 | +.31 | +.27 | -.21 | +.49 |
| | 4 | +.77 | +.59 | +.51 | +.39 | +.40 | +.63 |
| | 5 | +.88 | +.62 | +.22 | +.62 | +.14 | +.42 |
| | 6 | +.74 | +.22 | -.10 | +.28 | -.24 | +.50 |
| Veg. Hs. | 1 | +.86 | +.33 | -.34 | +.21 | -.46 | +.49 |
| | 2 | +.70 | +.15 | -.01 | -.19 | -.51 | +.76 |
| | 3 | +.55 | +.30 | -.05 | +.09 | -.16 | +.68 |
| | 4 | +.49 | +.09 | +.25 | +.06 | -.15 | +.72 |
| | 5 | +.83 | +.36 | +.23 | +.38 | +.20 | +.21 |
| | 6 | +.84 | +.06 | -.26 | +.25 | -.09 | +.77 |
| Grassy | 1 | +.41 | +.38 | +.36 | -.34 | -.38 | +.97 |
| | 2 | +.44 | +.41 | +.34 | -.36 | -.41 | +.97 |
| | 3 | +.83 | -.24 | -.38 | -.51 | -.67 | +.95 |
| | 4 | +.62 | +.27 | +.17 | -.25 | -.34 | +.96 |
| | 5 | +.43 | -.16 | -.19 | -.77 | -.78 | +.97 |
| | 6 | +.63 | +.50 | +.43 | +.03 | -.07 | +.96 |

sample standard deviations in the infrared bands does result in spectral separability. Standard deviations of grassy samples are typically twice as large as standard deviations of the other land use samples in either Band 6 or Band 7.

Coefficients of correlation were also investigated, but only one, r_{67} , of the six correlations proved to be helpful. The correlation between the two infrared bands was always +0.95 or greater (highly significant statistically) for the grassy samples. Conversely, the r_{67} for samples from other land uses were always +0.83 or less.

4.5 Further Analysis of Milwaukee and Chicago Subframes

4.51 Introduction

Overall classification accuracy for the Milwaukee subframe (reported in Section I) exceeded 90 per-cent, which was higher than the 87 per-cent reported in the Indianapolis subframe. But, the sole reason the former classification was more accurate than the latter rested upon the fact that the Milwaukee data were collected during the Summer months. Consequently, minimal confusion resulted between the "Grassy" and "Inner City" spectral classes, which caused most of the misclassification in the Indianapolis study.

Important lessons were learned from analysis of the Indianapolis subframe. Thus, further work was done on the Milwaukee and Chicago subframes.

4.52 Data Analysis

Using the qualitative results from the Indianapolis analysis, samples were re-chosen in Milwaukee County for a number of land uses, and the county was classified again (Figure 4.4). Gray-levels used for the display of the spectral classes are as follows:

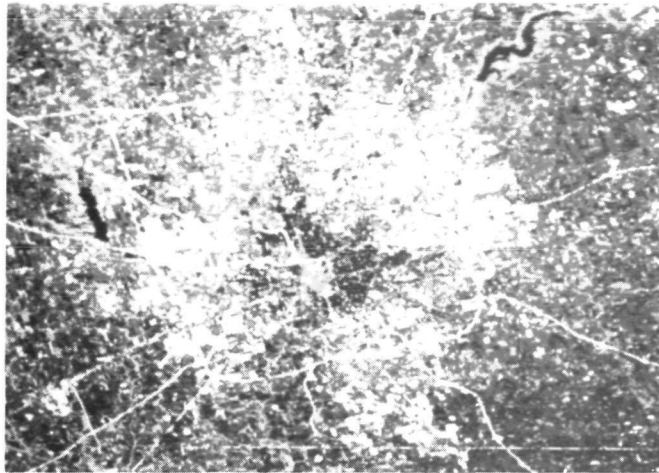


Figure 4.3 Computer-implemented land use classification
 of Marion County, Indiana.

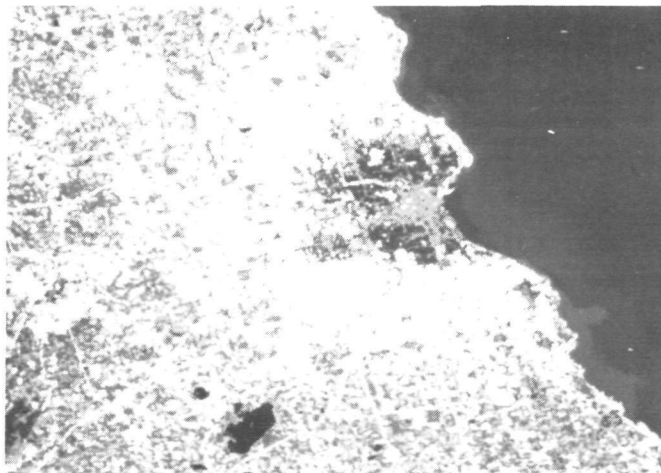


Figure 4.4 Computer-implemented land use classification
 of Milwaukee (second iteration).

| | | |
|-------------------|---|-----------------|
| Industry/Commerce | - | medium gray |
| Inner City | - | black |
| Suburban 1 | - | white |
| Suburban 2 | - | white |
| Wooded Suburban | - | light gray |
| Grassy | - | dark gray |
| Water 1 | - | dark gray |
| Water 2 | - | very dark gray |
| Water 3 | - | black |
| Water 4 | - | black |
| Water 5 | - | very light gray |
| Cloud | - | white |
| Shadow | - | black |
| Wooded | - | light gray |

The same principles of visual separation of classes apply to the image in Figure 4.4 as described for Figure 4.1. An additional visual separation must be made, however. "Wooded Suburb" and "Wooded" are both displayed as light gray. The former is located only within the built-up area of the county, while wooded is located in rural areas.

Four major changes were made in the classification. One of these was the combination of the "Road - downtown" and "Industry" classes. This resulted in a more extensive areal distribution of the Commercial/Industrial areas, more extensive than the two previous classes used and more accurate classification.

The second change is the addition of the "Wooded" spectral class. Wooded areas were accurately identified in the Indianapolis classification. Thusly, previous unsuccessful attempts at classifying wooded areas in Milwaukee were thought to be a result of inadequate training sets.

Thirdly, an attempt was made to classify newer, upper-income areas previously misclassified as grassy, located primarily in the suburbs of Brookfield, Elm Grove, and New Berlin. This was accomplished by creating a new suburban spectral class (both are displayed as white in the classification image). The resulting classification of suburban areas shows them to be more extensive than before. Newer, upper income areas were classified correctly, but country roads and areas of grassy (open or agricultural) land cover were also included.

The fourth change was simply to use one spectral class of grassy, instead of two.

4.53 Reclassification of Chicago Subframe

The new set of Milwaukee statistics was used to reclassify the Chicago subframe (Figure 4.5). Results were similar to the reclassification of Milwaukee. Commercial/Industrial areas were more accurately classified, as were newer, upper income areas.

4.6 Use of Geometrically Corrected ERTS Data

4.61 Introduction

ERTS data that has been "geometrically corrected" has been 1) deskewed, i.e., the five degree skew due to the earth's rotation beneath the satellite has been removed, 2) rotated, i.e., the thirteen degree tilt of the data due to the satellite's orbit has been removed, and the data is oriented North-South, and 3) scaled, i.e., computer printouts of the data are maps of the approximate scale of 1:24000.

Geometric correction of ERTS data is important for two reasons. One, the analyst can more easily work with the data. His ground truth will usually be in the form of North-South maps, or at least recorded onto North-South base maps. The ground truth can be easily transferred onto computer printouts which are oriented

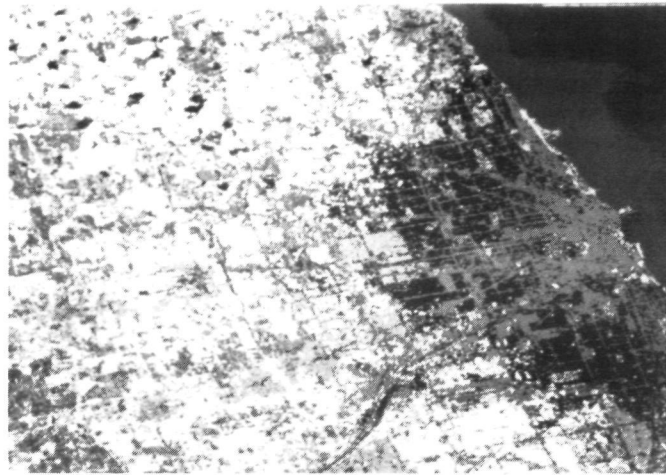


Figure 4.5 Computer-implemented land use classification of Chicago (second iteration).

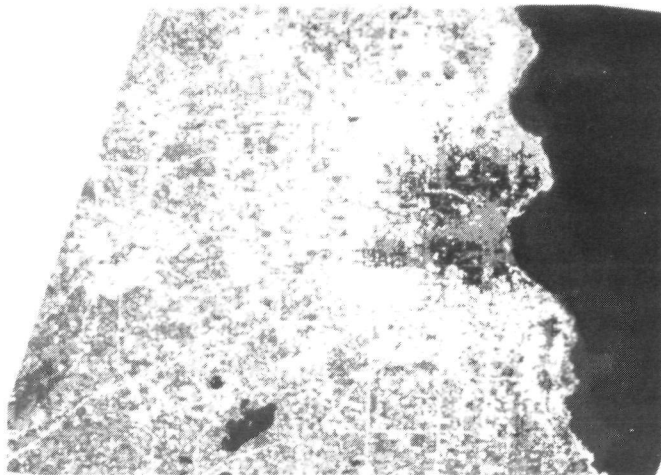


Figure 4.6 Computer-implemented land use classification of Milwaukee. Prior to classification, data was straightened (oriented north-south) and deskewed.

North-South and have constant scale both vertically and horizontally.

The second benefit of using geometrically corrected data is the resulting map which is presented to the "user" agency or individual. Similar to the benefits received by the analyst, the user will be able to locate himself more easily on geometrically corrected data. Also the user will probably have other areal data he has collected; he can more easily interface the non-ERTS data with geometrically corrected ERTS data.

4.62 Geometrically Corrected ERTS Classification Results

The Milwaukee statistics were used to classify Milwaukee County, using geometrically corrected data. Figure 4.6 shows the resulting classification, photographed from the digital display. Note that while the data has been oriented North-South and has been deskewed, the scale distortion due to the digital display format (square versus rectangular display of data points) has been introduced, i.e., horizontal distance is 20 per-cent greater than it should be. This problem is by-passed in Figure 4.7, which shows a portion of the Milwaukee classification printout overlaid with a U.S.G.S. 7.5' topographic quadrangle.

The Indianapolis data were also geometrically corrected, and subsequently classified (Figure 4.8).

Examination of both classification printouts -- Milwaukee and Indianapolis -- revealed a ± 2 per-cent error in the geometric correction, both North and South. If one overlays a classification printout onto a composite U.S.G.S. 1:24,000 quadrangle (such as the "Milwaukee and Vicinity" composite quadrangle), and lines up an areal feature with both maps, moving fifteen miles either North or South to another areal feature will reveal that the printout is misregistered five or six lines/columns. Notwithstanding, the error is minimal for the purpose of the great majority of analyses.



Figure 4.7 Computer-implemented Milwaukee area land use classification printout (geometrically corrected) overlaid with a U.S.G.S. 7.5' topographic quadrangle. Symbols: l=commerce/industry; M=inner city; O=suburban; -=grassy; (=wooded; I=wooded suburb; .=water; 1; S=water 2; X-water 3; +=water 4; L-water 5.

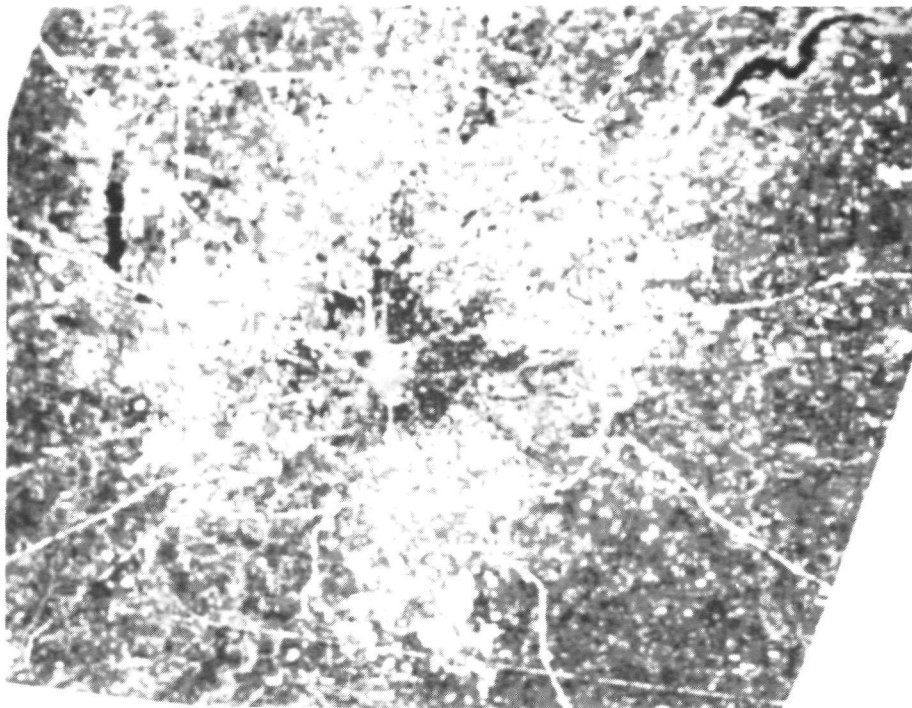


Figure 4.8 Computer-implemented land use classification of Indianapolis, Indiana (Marion County). Prior to classification, data was straightened (oriented north-south) and deskewed.

4.7 Use of Differing Histograms in Conjunction with the LARS Digital Image Display

4.71 Introduction

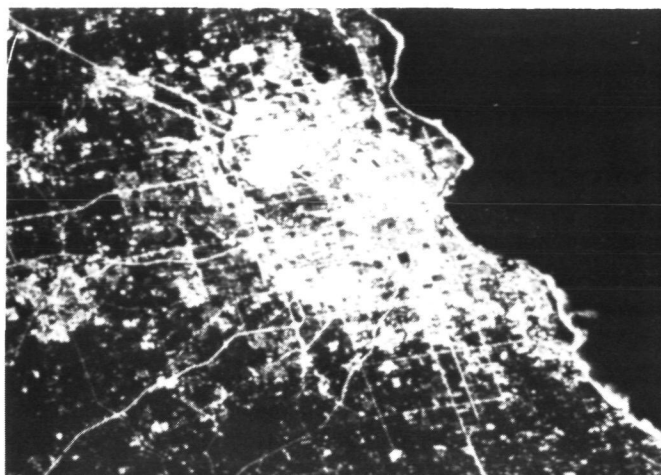
To begin an analysis of an ERTS test site, the researcher's most valuable tool is the imagery and functions of the digital display. Utilizing the lightpen for determining line/column coordinates of land marks and utilizing the enlargement function, he can save valuable time which would otherwise be spent making tedious measurements on computer printouts.

The digital display can display a maximum of sixteen gray-levels from black to white. Urban land uses have spectral responses ranging from 0 (usually water in Band 7) to over 100 (clouds in Band 6). Disregarding cloud spectral responses, the higher spectral responses in Band 6 (the ERTS band with the greatest data range) may get as high as 60.0. Consequently, a given set of display gray levels may not show all of the spectral variation in the ERTS data.

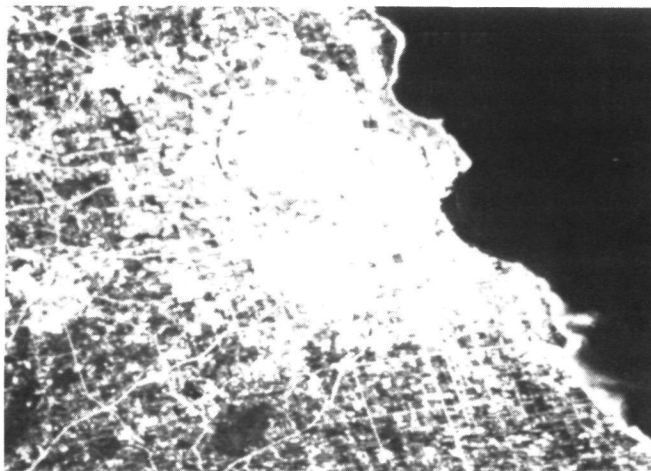
4.72 "Urban" versus "Rural" Histograms: The Example of the Milwaukee Subframe

At least two, and possibly three, different sets of histograms could feasibly be generated for use in viewing urban features on the digital display. Two different sets of histograms were generated for use in viewing urban features (within the Milwaukee subframe) on the digital display. One set of histograms was termed "urban"; the area selected for histogramming was in central Milwaukee. The other set of histograms was termed rural; histogramming was done outside of Milwaukee, in an agricultural area.

Figure 4.9 shows gray-scale imagery of two ERTS bands (Band 5, 0.6 - 0.7 μm ; Band 7, 0.8 - 1.1 μm). Differences



A

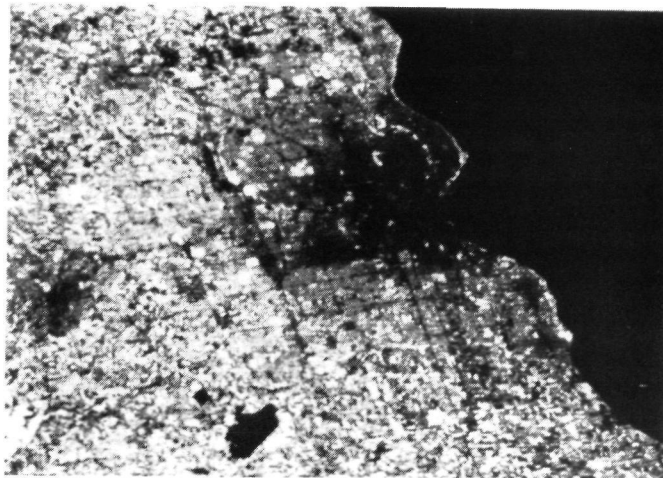


B

Figure 4.9 Grayscale imagery of Milwaukee County area. A and B are from the visible portion of the spectrum (Band 5, 0.6-0.7 μm), C and D are from the infrared (Band 7, 0.8-1.1 μm). A and C were from "urban" histograms; B and D from rural area histograms.



C



D

Figure 4.9 continued.

between the two histogram sets are indeed striking. Referring back to the classification image (Figure 4.6), the urbanized area is more correctly outlined in the Rural Band 5 image; the Urban Band 5 image excludes newer, upper income areas from the greater, typically bright urbanized complex. Inner City is clearly shown on the Rural Band 7 image, but is difficult to visually distinguish on either of the Urban images. Industrial areas are not evident on either of the two Rural images, yet are seen in the Urban Band 7 image.

4.8 Analysis of Gary, Indiana Area Subframe

4.81 Data Processing

Approximately half of an ERTS frame collected over the Southern part of Lake Michigan on October 1, 1972, is shown in Figure 4.10. The study area (outlined in the center) includes diverse land use types, including concentration of industry in Gary, Hammond, Whiting, and East Chicago, large residential areas, agricultural land, and forested areas.

Initially, the four ERTS bands were examined on a digital imaging display. Two of the images are shown in Figure 4.11. Several rectangular areas were selected for cluster analysis. The resulting cluster maps were used to pick small, rectangular areas for each desired land use class. This set of training samples was used to classify the entire study area, using a Gaussian maximum likelihood classifier.

4.82 Classification Results

The resulting classification is shown in conjunction with the raw data (Figure 4.11), and as Figure 4.12. Important features and land use classes are located with letters or numbers in Figure 4.12 and are also listed in Table 4.3. Gray levels

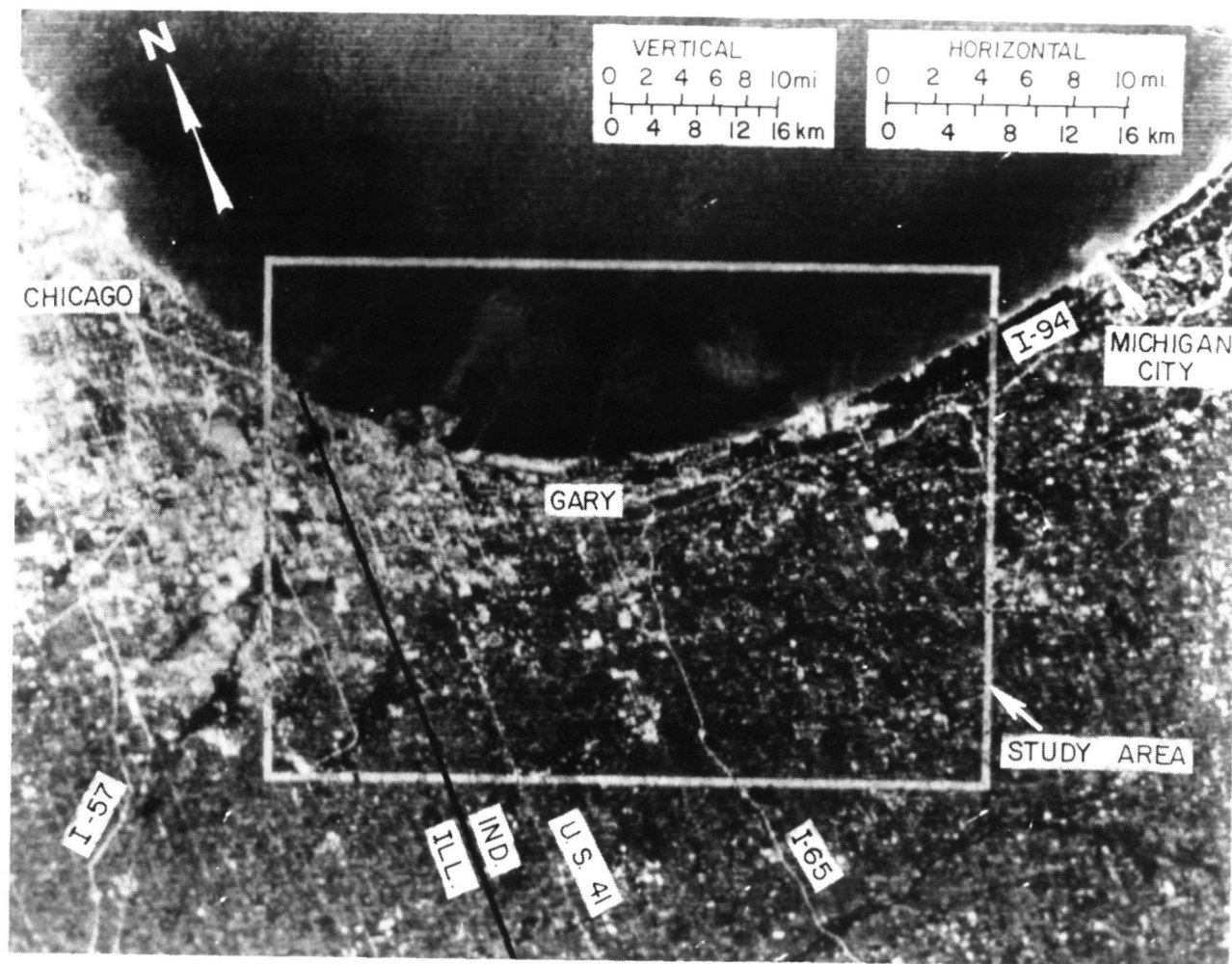


Figure 4.10 Photo from digital display of Gary, Indiana area showing location of the study area (outlined). Image is from Band 4 (0.5-0.6 μ m).

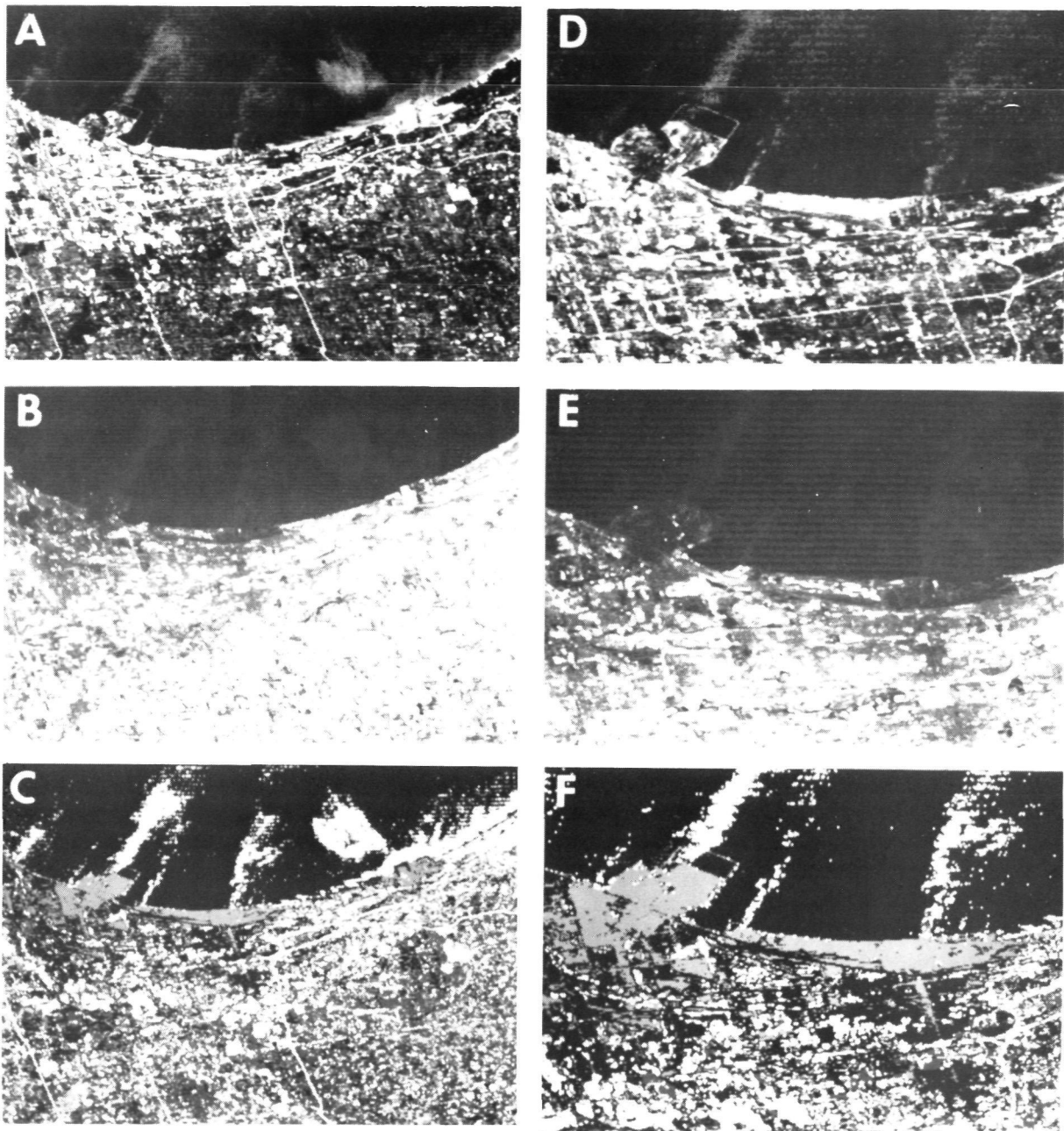


Figure 4.11 Photos from digital display, showing relationship between gray scale imagery and land use classification for Gary, Indiana. Image in A is from the visible portion of the spectrum (Band 4, $0.5-0.6\mu\text{m}$); B is from the reflective infrared (Band 6, $0.7-0.8\mu\text{m}$); C is a computer-implemented classification of the study area (see text for explanation of gray levels). Images in A, B, and C show the entire study area; enlargements of the northwestern portions of those three images are shown in D, E, and F, respectively. Horizontal length of A, B, and C is 54 kilometers (29 miles). Horizontal length of D, E, and F is 27 kilometers (17 miles); vertical length is 23 kilometers (14.5 miles). The true north-south line is rotated about 18 degrees counterclockwise to vertical. Horizontal scale is approximately three-fourths that of the vertical scale.



Figure 4.12 Photo from digital display of computer-implemented land use classification of Gary-Hammond area (see text for explanation of gray levels). Letters/numbers refer to feature of interest, a listing of which is found in Table 4.3 .

used for the display of the spectral classes are as follows:

| | |
|-----------------------------|-------------|
| Industrial/Commercial | medium gray |
| Older Housing | black |
| New Housing | white |
| Trees | light gray |
| Grassy (open, agricultural) | dark gray |
| Water | black |
| Smoke | white |

Older housing and water have been assigned the same gray level (black), but consideration of their areal distributions prevents confusion between the two. Water is largely restricted to Lake Michigan and to several other water bodies, such as Wolf Lake. Older housing is located between coastal industrial establishments and newer housing. Smoke and newer housing also have the same gray levels (white). Smoke, however, is found only over Lake Michigan, while newer housing is located to the South.

Agricultural areas (shown as dark gray) were identified in the Southern part of the study area. This class included cropland, pasture, and idle land in rural areas, as well as parks, golf courses, and open land in urban areas. Wooded areas (shown as light gray) are commonly associated with the drainage pattern of the study area. Three principal stands of trees appeared, in conjunction with the Little Calumet River, Deep River, and the dunes park area along Lake Michigan. Water, displayed as black, is located in Lake Michigan and other, smaller water bodies such as Wolf Lake.

Newer housing developments, shown as white, are located on the fringes of the urbanized area, in the municipalities of Munster, Highland, Griffith, and Merrillville. The majority of the structures were built prior to World War II. Lawns (grass) and streets are the two primary constituents of this spectral

Table 4.3 Features of Interest, Land Uses, and Major Highways
Indicated in Figure 4.12

| L | Feature, Land Use, or Highway ² | Spectral Class ³ |
|---|--|-----------------------------|
| A | Smoke Plume | smoke |
| B | Inland Steel | commercial/industrial |
| C | United States Steel | commercial/industrial |
| D | Bethlehem Steel | commercial/industrial |
| E | Oil Refineries | commercial/industrial |
| F | Wolf Lake | water |
| G | Lake Michigan | water |
| H | Gary - Central Business District | commercial/industrial |
| J | Highland - subdivision | newer housing |
| K | Indiana Harbor | commercial/industrial |
| L | Port of Indiana | commercial/industrial |
| M | Gary Municipal Airport | newer housing |
| N | Indiana Dunes State Park | wooded |
| O | Agricultural Area | grassy/agricultural |
| P | Indiana - Illinois State Line | --- |
| Q | Wicker Memorial Park | grassy/agricultural |
| R | Trees Along Deep River | wooded |
| S | Hammond - Residential Area | older housing |
| T | East Chicago - Residential Area | older housing |
| U | Munster - Subdivision | newer housing |
| V | Gary - Residential Area | older housing |
| 2 | Interstate Highway 80-94 | newer housing |
| 3 | Interstate Highway 80-90 | newer housing |
| 4 | U. S. Highway 12 | newer housing |
| 5 | Interstate Highway 94 | newer housing |
| 6 | Illinois Highway 394 | newer housing |
| 7 | U. S. Highway 41 | newer housing |
| 8 | Interstate Highway 65 | newer housing |
| 9 | U. S. Highway 30 | newer housing |

class. Therefore, four-lane highways were also classified as newer housing.

Older residential, displayed as black, consists of areas developed prior to World War II. They are found in Hammond, Whiting, East Chicago, and Gary. Closely spaced rooftops, along with mature vegetation (large trees) are the reasons for the spectral separability of this class.

Industrial/Commercial areas (shown as medium gray) are usually void of vegetation. They are characterized by the occurrence of rooftops, parking lots, streets, and bare ground. Examples include Inland Steel, U.S. Steel, Standard Oil, Bethlehem Steel, the Gary Central Business District, and Broadway Plaza Shopping Center (Figure 4.12).

4.83 Industrial Land Use Classification

The land use classes identified in this study correspond well with the classes proposed by Anderson, Hardy, and Roach in the U. S. Geological Survey Circular 671, and also with those developed in previous ERTS urban analyses (4, 1, 5, 6). Further investigations were made, however, into industrial areas in this analysis because of their large areal extent in the Gary-Hammond area.

Five spectral classes of commercial/industrial land use were developed. Two of the classes are associated with closely spaced rooftops; the other three are associated with gravel or sandy areas in industrial areas, adjacent to the rooftop classes. The Northern part of the classification image is shown in Figure 4.13. Gray levels used for the classification image are as follows:

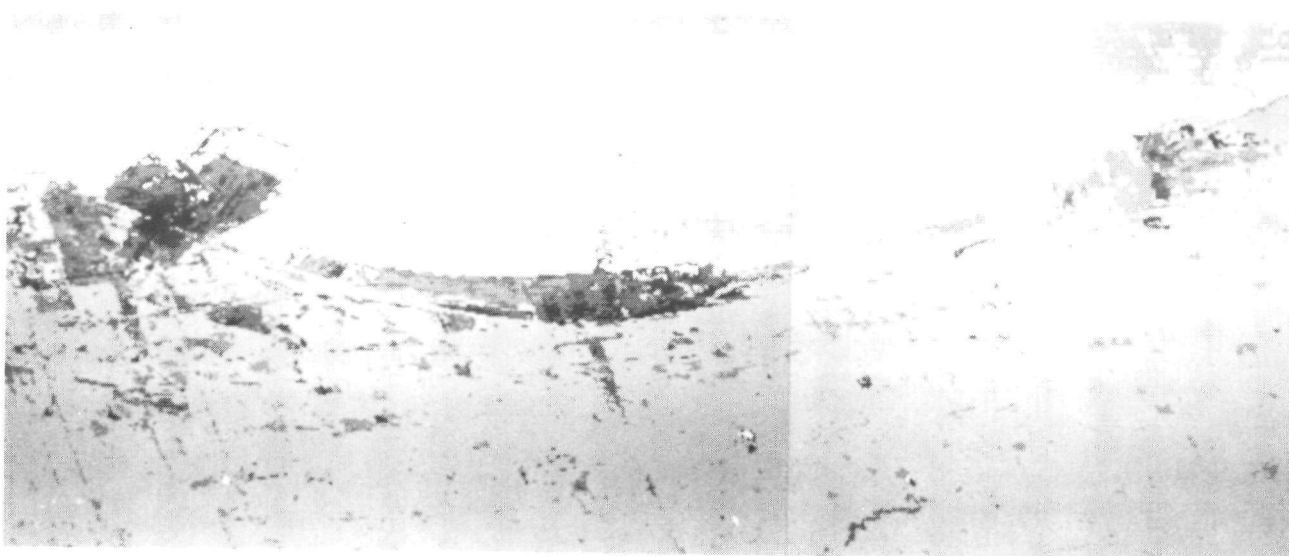


Figure 4.13 Photo from digital display of computer-implemented land use classification of Gary-Hammond area (northern part of study area) using gray levels which emphasize the industrial land uses. Class shown as black is dark roofing material; dark gray is lighter-colored roofing material; medium gray is gravel/sandy areas; smoke is white. All other spectral classes are shown as light gray.

Table 4.4. Classification Accuracy for Test Samples for Gary Site

| Land Use | Percentage of Data Points Classified As: | | | | | |
|---------------------|--|------------------|------------------|------------------|------------------|------------------|
| | C/I ¹ | OHg ² | NHg ³ | Wod ⁴ | A/G ⁵ | Wtr ⁶ |
| Commerce/Industry | <u>89.8</u> | 7.3 | 0.3 | ---- | 0.2 | 2.5 |
| Older Housing | 0.9 | <u>97.9</u> | 0.9 | 0.3 | ---- | ---- |
| Newer Housing | 0.6 | 4.0 | <u>94.0</u> | ---- | 1.4 | ---- |
| Wooded | ---- | 2.0 | ---- | <u>94.4</u> | 3.5 | ---- |
| Agricultural/Grassy | 0.8 | 27.4 | 3.2 | 3.1 | <u>65.5</u> | ---- |
| Water | 0.8 | ---- | ---- | ---- | ---- | <u>99.2</u> |

X Classification Accuracy by Class = 90.3%.

¹Commerce/Industry

²Older Housing

³Newer Housing

⁴Wooded

⁵Agricultural/Grassy

⁶Water

| | |
|--------------------------------|-------------|
| Rooftops (dark reflectance) | black |
| Rooftops (bright reflectance) | dark gray |
| Gravel/Sandy areas (3 classes) | medium gray |
| Smoke | white |
| All Other Classes | light gray |

The reader will note that the same classification is shown in both Figure 4.12 and Figure 4.13; the only difference is the assignment of gray levels to the spectral classes.

The class shown as black in Figure 4.13 is associated primarily with dark roofing material, but also with large coal piles. Large areas of this spectral class are associated with the three large steel firms in the study area -- Inland, U. S., and Bethlehem. The other rooftop class, displayed as dark gray, is associated with brighter reflecting rooftops. Reasons for the three spectral categories of gravel/sandy areas (all displayed as medium gray) are not entirely clear at this time, but they probably relate to both the color of the material and the presence/lack of sparse vegetative cover. Two large areas of gravel/sandy material are located in the Northwest part of the study area, one between the large building complexes of Inland and U. S. Steel companies and the second in the large oil refining district in the Whiting-East Chicago area. In the Northeastern part of the study area, another large area of gravel/sand is located West of the buildings of Bethlehem Steel. This area was dominated by one of the three classes of gravel/sand, and had a particularly high spectral reflectance in both the visible and infrared portions of the spectrum. The ground cover in this area is the dune sand typical of this locale.

The final class used in the classification scheme, the white areas in Lake Michigan, is speculated to be smoke coming from coastal industrial establishments. Spectrally, the class

is similar to water, having a very dark reflectance in the infrared (Figure 4.11-B). Several facts, however, when considered as a whole, lead one to conclude it is smoke. The linear, parallel arrangements of the data points, extending some 30 miles into Lake Michigan, are contrary to the circulation patterns in the lake. Moreover, meteorological records report that the wind was out of the South-west on the morning of the ERTS pass.

While smoke was probably identified in the Western part of the study area, the smoke data points along the coast in the East were probably water. Moreover, the large area classified as smoke Northwest of Bethlehem Steel was probably a thin cloud. Despite the spectral confusion in these areas, the partial separability does warrant further investigation of the phenomena of smoke located over water bodies.

4.84 Classification Accuracy

An attempt was made to determine the classification accuracy by a sampling method. A number of rectangular test areas were determined for each land use, and the class accuracy determined (Table 4.4). Water, wooded areas, older housing, and newer housing were all identified with over 90 per-cent accuracy. Trouble was encountered in industrial/commercial areas, most of the misclassification being attributed to older housing. The poorest classification accuracy was in grassy and agricultural areas, where less than 70 per-cent of the data points were accurately classified. Misclassification of these areas was of two major types. One, areas in agricultural regions associated with darker colored soils proved difficult to separate from older housing. One such area was located South of Little Calumet River, in Munster and Highland. The other type of misclassification was in undeveloped

Table 4.5. Land Use Area Calculations for Study Area
(excluding Lake Michigan)

| Land Use | Number of Data Pts. | Number of Acres | Number of Hectares | % of Study Area |
|----------------------------------|------------------------|--------------------|-----------------------|--------------------|
| Commerce/Industry ¹ | 25766 | 28343 | 11479 | 8.2 |
| Older Housing ¹ | 56528 | 62181 | 25183 | 18.0 |
| Newer Housing | 28540 | 31394 | 12714 | 9.1 |
| Wooded | 52346 | 57581 | 23320 | 16.6 |
| Agricultural/Grassy ¹ | 150982 | 166080 | 67262 | 48.0 |
| Water | 499 | 549 | 222 | 0.2 |
| TOTAL | 314661 | 346127 | 140181 | 100.0 |

¹Adjustments made in accordance with test classification accuracy (see Table 2).

marshland adjacent to industrial areas. Large areas South of U. S. Steel and along U. S. Highway 12 were misclassified as older housing.

4.85 Area Calculation

Important tabular data can be generated from the machine processing of ERTS data. Table 4.5 contains an estimate of the proportion of the study area (excluding Lake Michigan) allocated to the various land uses, obtained by a simple tallying of the numbers of data points in each land use. Adjustments were made for agricultural/grassy areas, commercial/industrial, and older housing, relative to the misclassification between these three land uses. Acreages were obtained by multiplying the number of data points by 1.1, the approximate pixel area of ERTS. The data in Table 4.5 could have been reported by smaller areal units, such as municipalities, townships, or census tracts, by storing the desired boundaries in the computer.

4.9 Conclusions

The results of these investigations suggest that computer analysis of ERTS MSS data may be a valuable tool for the urban-regional planner. Although only gross land use inventories may be made, because of the satellite's resolution, timely updating of a metropolitan area's data bank would be invaluable. Detection of land use change by the satellite would indicate where detailed studies (either by aerial photography or direct field investigation) ought to be pursued. Such detection would have been possible before costly air photo coverage or ground observations of the entire area were made.

At the state/regional or national levels, machine processing of ERTS data may be totally adequate. State officials are not disinterested in local land use problems, but they are concerned more with broad trends within a large area. Classification accuracies of 87 to 92 per-cent may be adequate for their purposes.

5.0 Water Resources Research

5.1 Introduction

In the state of Indiana most of the surface-water is stored in man-made lakes or reservoirs. These reservoirs are primarily used for flood control, water pollution abatement, and recreation. However, in some cases their content is also used for municipal and industrial water supply. Acquisition of hydrological data such as surface area of lakes and water quality is necessary for adequate planning and managing of water resources.

Conventional data collection techniques have the disadvantage of requiring a great deal of time and effort, and usually they do not give a comprehensive picture of the actual situation. Therefore, new techniques must be developed and evaluated. At the present time, remotely sensed data from aircraft and spacecraft altitudes are satisfactorily being utilized in several disciplines as a means of gathering useful information about man's environment. It is, then, the purpose of this investigation to assess the feasibility of using ERTS-1 MSS data and the LARSYS computer-aided processing and analysis techniques to obtain information useful for more effective management of our water resources.

5.11 Objectives

The specific objectives of this research were:

- (1) To evaluate the utility of the multispectral data obtained from the ERTS-1 MSS sensor system for mapping spectral variations in water bodies.
- (2) To compare these spectral variations with the spectral response obtained from aircraft altitudes in order to examine possible loss of spectral information in water bodies from satellite altitudes.

- (3) To determine the accuracy with which acreages of water bodies can be estimated from ERTS-1 MSS digital data and computer-aided processing and analysis techniques.

5.2 Multispectral Classification of Lakes Freeman and Shafer from ERTS-1 MSS Data

Multispectral scanner data collected by the ERTS-1 satellite over Northern Indiana on May 4, 1973 were analyzed. Six spectral groups were defined initially, using *CLUSTER. The resulting statistics were then utilized to define the training classes for the supervised classifier (*CLASSIFYPOINTS). The six spectral classes in the initial classification corresponded to:

- (1) Lake Shafer water
- (2) Lake Freeman water
- (3) Banks (edge of water bodies)
- (4) Agricultural
- (5) Forest
- (6) Soils

The spectral characteristics of the waters from Lake Shafer and Lake Freeman are very similar, as indicated by a separability¹ or divergence value of 146 between these two water bodies. This is a very low value of separability. We have therefore concluded that the waters of both lakes have spectral responses that are very similar and should be defined as a single spectral class. Thus, the final classification contained the following spectral classes:

1 Swain and King (1973) have reported on the relationship between the separability values and percent correct classification. For a transformed divergence value of 146 the percent correct classification would be just over 50%.

- (1) Lake water (Lakes Shafer and Freeman)
- (2) Water edge (a mixture of water and vegetation)
- (3) Agricultural
- (4) Forest
- (5) Soils

In order to determine if the lack of more than one spectral class of water found in Lakes Shafer and Freeman was caused by a loss of spectral information due to the altitude of the ERTS-1 satellite, supporting aircraft MSS data (low altitude) was then analyzed.

5.3 Multispectral Classification of Lake Freeman from Aircraft MSS Data

Multispectral scanner data were gathered in 12 bands by the ERIM* scanner system. These data were collected over Lake Freeman approximately an hour after the ERTS-1 overpass on May 4, 1973 at an altitude of 10,000 feet.

The classification was performed using training fields selected on the basis of photointerpretation of color, color IR and B & W photography taken simultaneously with the scanner data. The set of training classes included forests, crops and soils. However, in order to define water spectral classes, it was necessary to perform a non-supervised classification of water only. Thus, seven spectrally separable groups of water were defined. The coincident spectral plot of the seven water classes, the forest, crops and soils is shown in Figure 5.1.

It should be noted that the seven groups of water have an unusual spectral response; that is, in every reflective band the seven classes of water have a response that ranges from high to low; they do not show a complicated signature as a function of

* Environmental Research Institute of Michigan



wavelength. However, it is interesting to observe in Figure 5.1 that this is not the case with the thermal band (9.3-11.7 μm). In fact, there seems to be no difference in radiant temperatures among the seven classes of water.

The results of the classification are shown in Figure 5.2, in which only the seven water classes have been displayed using different grayscale symbols. Note that the classes of water in Figure 5.2 show a particular spatial distribution. They are distributed from East to West in parallel bands showing the brightest classes on the East side of the lake and the darkest on the West side. This pattern is unlikely due to either depth effects or water quality (turbidity). A close inspection of the photography taken simultaneously with the scanner data showed that the glare (specular reflection) from the water surface had an intensity distribution similar to the pattern shown in the classification. This suggests that the seven spectral classes of water defined by the clustering algorithm defined different intensities of specular reflection, which are a function of the sun-scanner-look angle.

Figure 5.3 shows the CRT (cathode ray tube) imagery of Lake Freeman in the 0.52 - 0.57 μm band. These data were collected from an altitude of 10,000 feet and at approximately 10:00 local time. It is interesting to note in Figure 5.3 that the West side of the lake appears light in tone and dark on the opposite side. It also shows the intense specular reflection from the water surface on the East side of the lake. The glare intensity decreases as one moves from the East side towards the center of the scene. The evidence of this monotonic decrease in specular reflection can be seen in Figure 5.4 which shows the graph of a scan line across the lake in three different portions

* Environmental Research Institute of Michigan

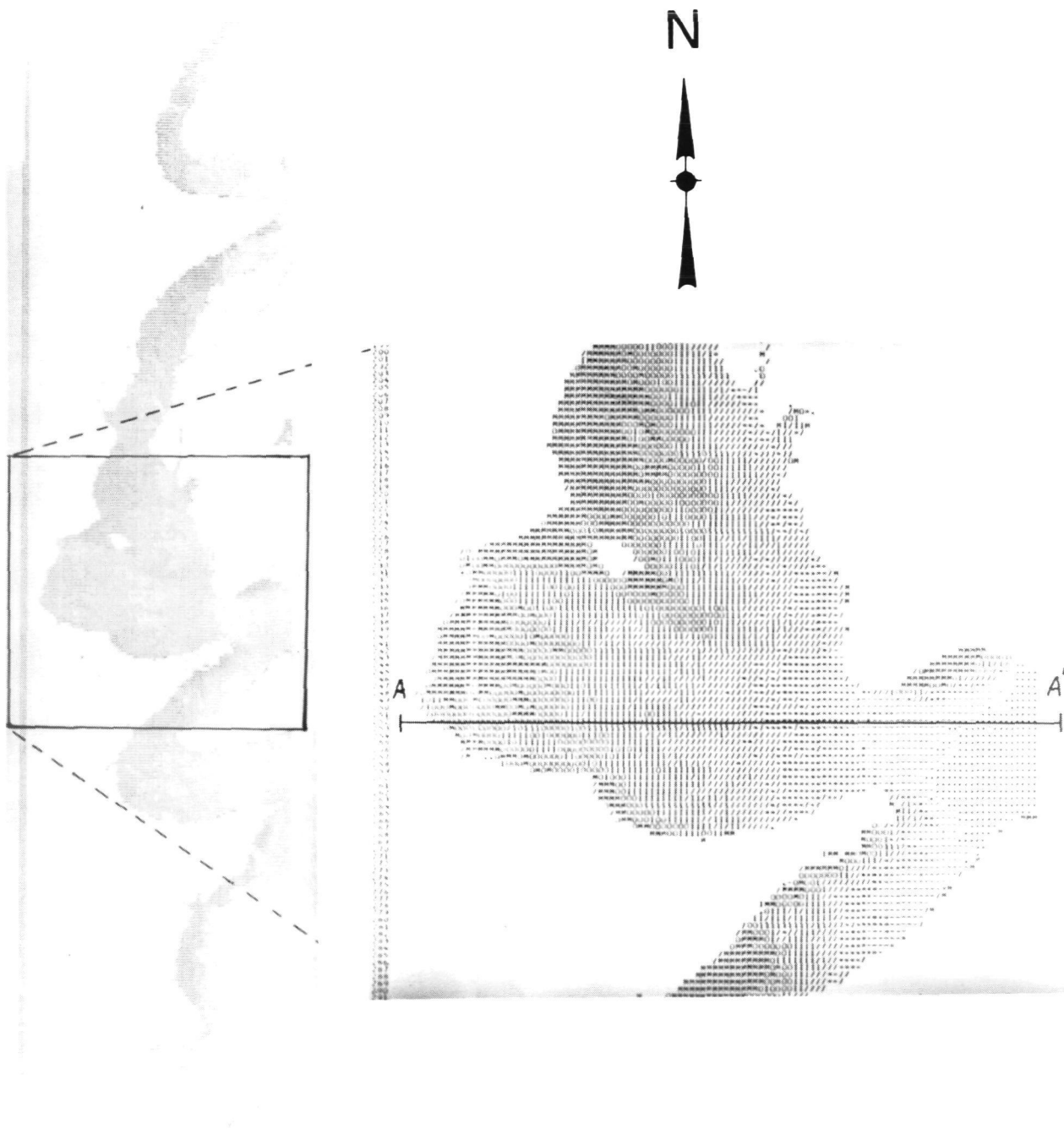


Figure 5.2 Seven spectrally separable classes of water in Lake Freeman, Indiana from Michigan scanner data.



Figure 5.3 CRT imagery from 10,000 feet altitude in 0.52 - 0.57 μm band. Note the sun-scan look angle effect on water and other cover types. Water appears dark on the side where everything else appears light.

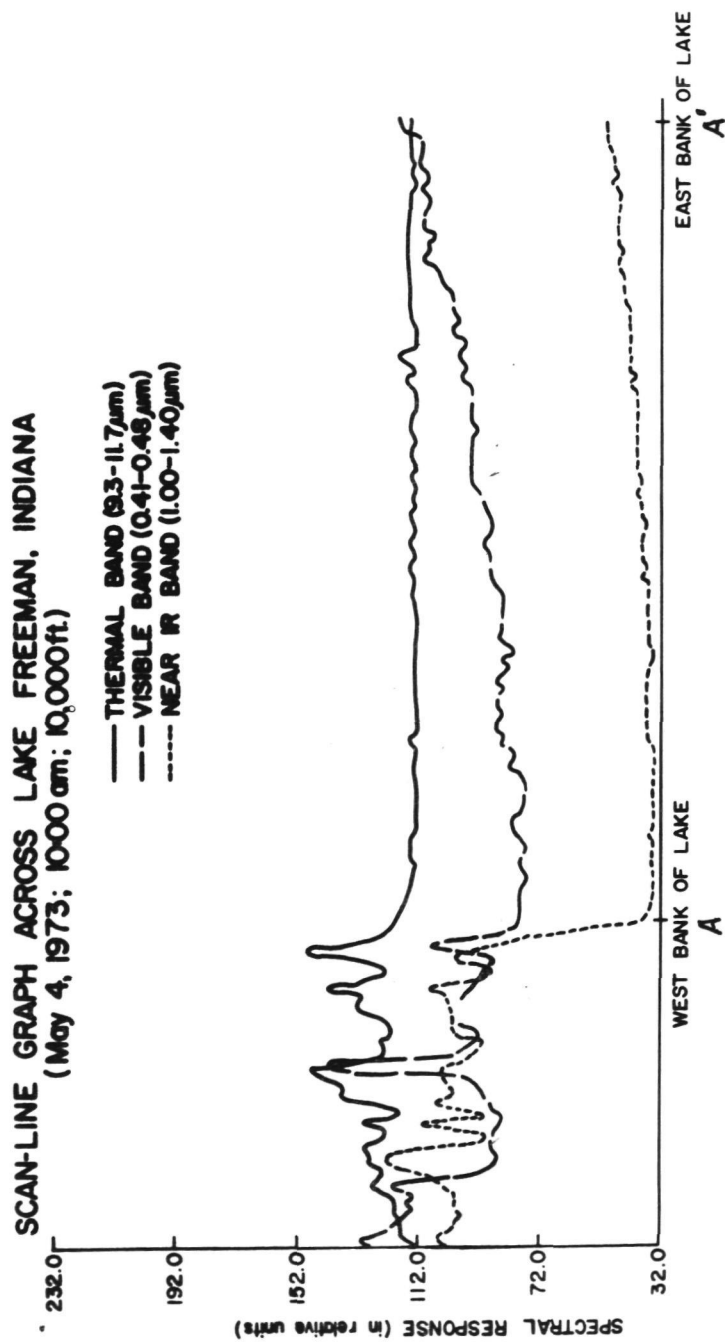


Figure 5.4 Spectral response graph across Lake Freeman. The location of the traverse A-A' is illustrated in Figure III-3.

of the spectrum. The location of traverse A-A' is illustrated in Figure 5.2. Note that the intensity of the specular reflection is larger in the visible (0.41 - 0.48 μm band) than in the near infrared (1.00 - 1.40 μm band) region of the spectrum. On the other hand, the thermal infrared (9.3 - 11.7 μm band) response does not seem to be affected by the scanner look angle. In short, the scanner look angle effect is important in the visible region of the spectrum, less important in the reflective infrared, and it does not affect the thermal response. The lighter side of the scene shown in Figure 5.3 is caused by the scattering of the incoming solar radiation by the atmospheric constituents and it has been found that this effect is not as pronounced in the near infrared wavelengths as is the case with the shorter wavelength.

5.4 Discussion of Satellite and Aircraft Results

Thus far, we have described the multispectral classifications of Lake Freeman utilizing MSS data from both spacecraft and aircraft altitudes. As previously stated, these two sets of data were collected over the same target at approximately the same time, and one of the objectives of this research was to compare the multispectral classifications from ERTS-1 and supporting aircraft data. The results described in the previous sections indicate that there are some advantages and disadvantages associated with each set of data. It is clear that one of the major drawbacks inherent in ERTS-1 data is its coarse spatial resolution. However, ERTS has the advantage of producing an almost orthogonal image in which scanner-look angle effects are minimal and can be disregarded.

In the case of aircraft data, the effects of the sun-scanner-look angle are so pronounced that under certain circumstances the data is useless for surface-water studies because

any spectral characteristic due to depth, turbidity or any other water-quality parameter could be completely masked by the strong specular reflections from the water surface. Because this problem (specular reflection) was encountered in the aircraft data from Lake Freeman, it was not possible to establish if, indeed, the presence of only one spectral class of water in Lakes Shafer and Freeman, as established from the ERTS-1 data, was due to a loss in spectral information caused by the altitude of the spacecraft. However, previous work with ERTS-1 data collected over Lake Texoma² had indicated that it was possible to separate several distinct spectral classes of water within the same lake.

Therefore, several other lakes in Northern Indiana were analyzed to determine if there were other spectral classes of water besides the one found in Lakes Shafer and Freeman. The resulting classification showed that five separable spectral classes of water were present in this Northern Indiana region. Some of the lakes fell under one spectral class and the others under one of the largest reservoirs, that is, the Salamonie Reservoir contained all five spectral classes.

5.5 Multispectral Classification of the Salamonie Reservoir from ERTS-1 Data

The same data set and procedures of analysis followed for the classification of Lake Freeman and Lake Shafer were utilized for the study of other lakes in Northern Indiana. The results showed that the spectral characteristics of the Salamonie Reservoir were different from those of Lakes Freeman and Shafer. In addition, it was possible to separate and map five distinct categories of water within the Salamonie Reservoir, as illustrated in Figure 5.5.

2 The results of this research have been reported in the first ERTS-1 Symposium Proceedings, X-650-73-10, Goddard Space Flight Center, September 1972.

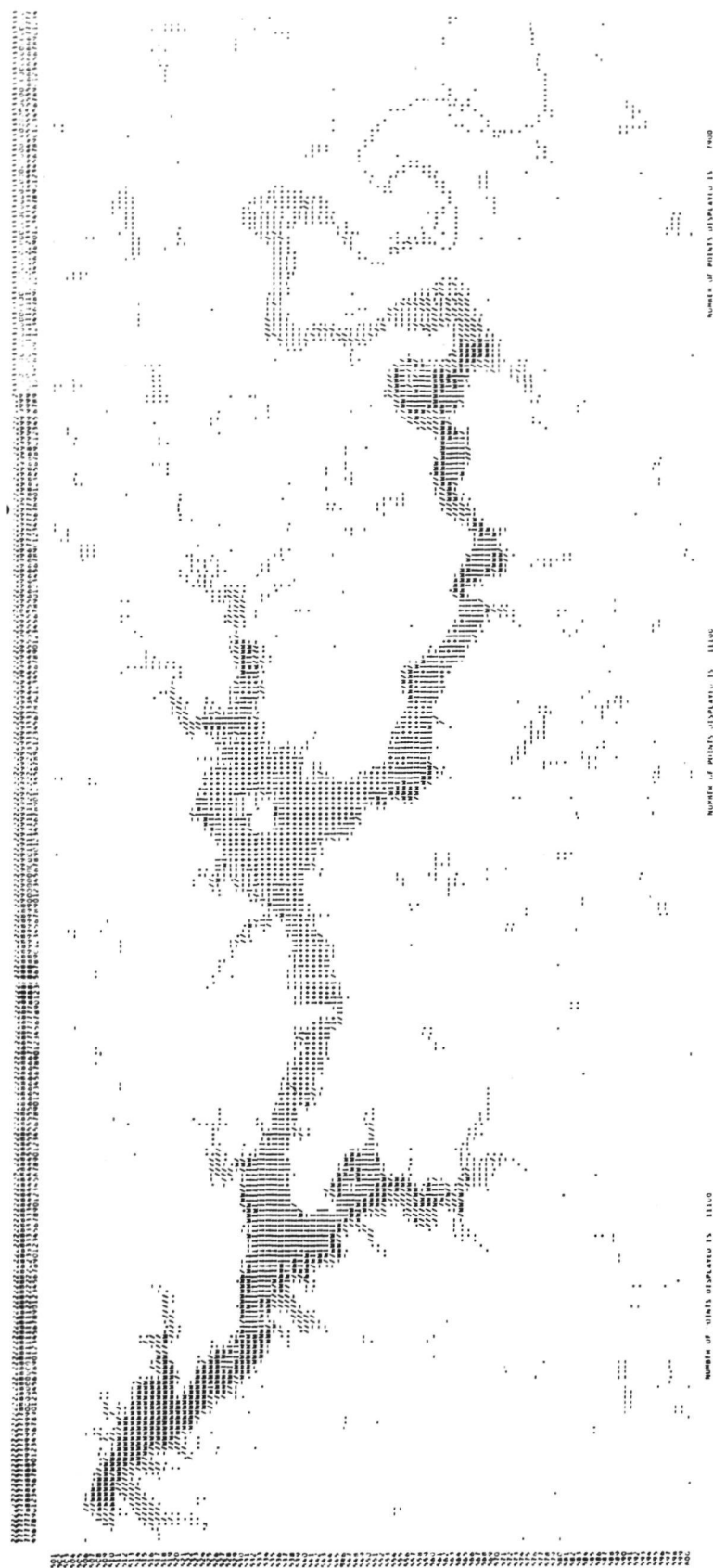


Figure 5.5 Multispectral classification map of the Salamonie Reservoir in North-eastern Indiana.

Table 5.1 shows the spectral response of the five classes of water in the four ERTS-1 MSS bands.

Table 5.1

Mean Response of Spectral Classes of Water
in the Salamonie Reservoir

| | 4 | 5 | 6 | 7 |
|---------------|-------|------|------|------|
| Water Class A | 33.4* | 27.4 | 19.3 | 6.2 |
| Water Class B | 40.0 | 36.2 | 22.4 | 5.7 |
| Water Class C | 42.5 | 41.1 | 25.0 | 6.0 |
| Water Class D | 45.1 | 46.3 | 30.0 | 7.0 |
| Water Class E | 31.4 | 28.0 | 31.0 | 10.2 |
| Edge Class | 27.4 | 20.8 | 25.5 | 12.3 |

A close look into the location and distribution of the "Water Class E" displayed as a dash (-) in Figure 5.5 and a thorough inspection of the low altitude photography of the area indicate that this particular spectral class occurs in the shallow areas of the reservoir where trees have been partially covered with water. The same class also occurs outside of the reservoir in areas where water has ponded in agricultural fields. Therefore, it appears that this spectral class represents a mixture of cover types, but is dominated by the spectral response of the water.

The other water classes, that is A, B, C, and D, are believed to indicate different levels of turbidity, as suggested

* These values refer to the mean relative response from the ERTS-1 scanner system. They range from 0 to 128 for bands 4, 5, and 6. Band 7 has a dynamic range from 0 to 64. The standard deviation for these mean relative responses is within plus or minus one relative unit.

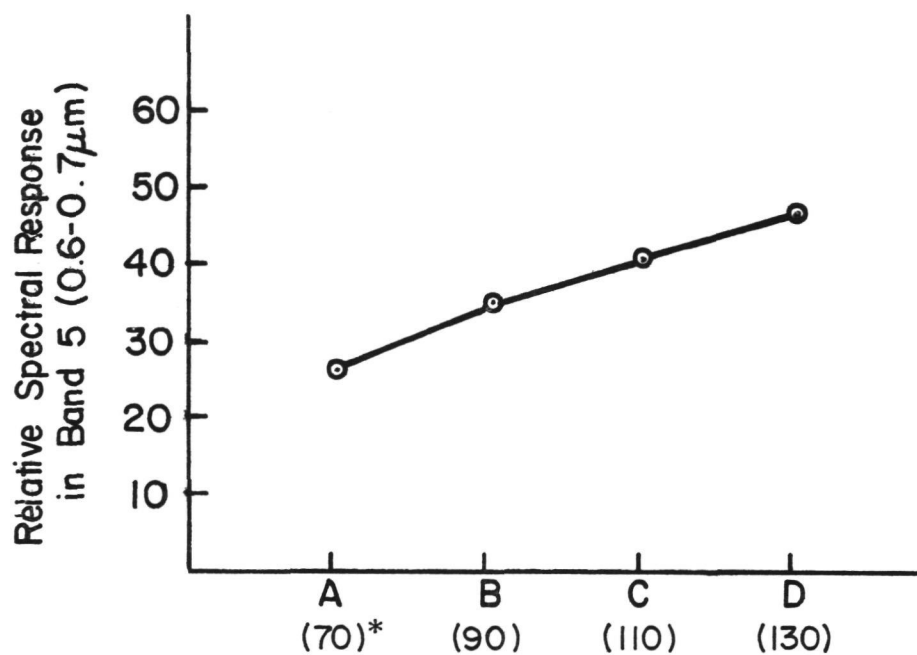
by the continuous increase in spectral response (shown in Table 5.1) in the visible channels, particularly in band 5 (0.6 - 0.7 μm). Weisblatt et al, (1973) have reported that an increase in the level of turbidity in water bodies causes a linear increase in the spectral response in the visible region of the spectrum, especially in band 5 (0.6 - 0.7 μm) of ERTS-1. Their results have been corroborated by similar results obtained from field work at LARS with an EXOTECH-100 spectroradiometer. The ERTS-1 data used by Weisblatt and co-investigators were collected on May 8, 1973, and the data utilized in this study were gathered on May 4, 1973. A graph of the spectral response of the different classes of water in band 5 obtained in our study is shown in Figure 5.6. The numbers in parentheses represent approximate levels of turbidity in parts per million (ppm), as shown by Weisblatt et al.

Thus, it is evident that one can detect differences in spectral response within a water body using ERTS-1 data. In addition, the distinct spectral classes of water found in the Salamonie Reservoir may indicate different levels of turbidity. However, further work and more reliable surface observations are needed to determine in a more quantitative manner the correlation between the water spectral classes obtained from ERTS-1 MSS data and the amount of suspended solids present in the water.

The next phase of this investigation was to determine the accuracy with which acreages of water bodies can be estimated from ERTS-1 MSS data.

5.6 Water Acreage Estimation from ERTS-1 MSS Data

Seventeen lakes and reservoirs in Northern Indiana were selected to determine the accuracy with which acreages of surface-water can be estimated from ERTS-1 multispectral data



*Spectral classes of water. The numbers in parenthesis represent approximate levels of turbidity in ppm (Adapted from Weisblatt et. al., 1973).

Figure 5.6 Relative spectral response in Band 5 of the water classes in the Salamonie Reservoir.

and using the LARSYS processing and analysis techniques. The seventeen lakes ranged in size from 15 acres up to 1864 acres. Their names, locations and size are given in Table 5.2.

The acreages shown in Table 5.2 were obtained from data published by the U. S. Department of the Interior Geological Survey in cooperation with the Indiana Department of Natural Resources. These acreages have been defined by USGS, based upon "established" water levels. For many years, records of the water-surface elevations of many lakes in Indiana have been collected by the Geological Survey. The established level is that elevation set by the courts to which the average level of the lakes is to be held. It is always set as the average level that has prevailed for a number of years. The surface area of a particular water body that corresponds to the "established level", is therefore the defined acreage of the reservoir.

Comparison of these average levels with the levels measured during the course of three years (1969, 1970 and 1971) indicates that the levels for the month of May are closely represented by the averaged levels (surface area) shown in Table 5.2.

Thus, a frame of ERTS-1 multispectral scanner data covering the Northern part of Indiana and collected on May 4, 1973 was selected in order to pursue the objective of determining the accuracy with which acreages of water bodies could be estimated utilizing digitized satellite data in conjunction with computer-aided classification techniques. The scene ID of the frame is 1285-15592, and its corresponding LARS run number is 73051600.

Table 5.2 Mean Areas of Lakes in Indiana Averaged over three years (1969,1970,1971)

| | <u>LAKE</u> | <u>COUNTY</u> | <u>SURFACE AREA (Acres)</u> |
|----|--------------|---------------|---------------------------------|
| 1 | Bass | Starke | 1400 |
| 2 | Maximkuckee | Marshall | 1864 |
| 3 | Bruce | Pulaski | 245 |
| 4 | Muskelonge | Kosciusko | 32 |
| 5 | Fish | Kosciusko | 15 |
| 6 | Yellow Creek | Kosciusko | 151 |
| 7 | Beaver Dam | Kosciusko | 146 |
| 8 | Loon | Kosciusko | 40 |
| 9 | Caldwell | Kosciusko | 45 |
| 10 | Silver | Kosciusko | 102 |
| 11 | Rock | Kosciusko | 56 |
| 12 | Langenbaum | Starke | 48 |
| 13 | Carr | Kosciusko | 79 |
| 14 | Nyona | Fulton | 104 |
| 15 | South Mud | Fulton | 94 |
| 16 | Zink | Fulton | 19 |
| 17 | Hartz | Starke | 28 |

Using the four available spectral bands of ERTS-1, these data were classified into five spectral groups including the following classes:

- (1) Water
- (2) Water-edge
- (3) Agricultural
- (4) Forest
- (5) Soils

Acreages were estimated by multiplying the number of resolution elements classified as water times a conversion factor of 1.12 acres per resolution element. (This factor had been previously determined by analysis of several ERTS-1 data sets.)

The resulting acreages estimated from the ERTS-1 data for the seventeen lakes in Northern Indiana indicated that there was a consistent under-estimation of the size of the lakes. These results are shown in Table 5.3 where the lake numbers correspond to the seventeen lakes shown in Table 5.2.

Table 5.3

Initial Acreage Estimates from ERTS-1 Data

| <u>Lake Number</u> | <u>USGS Data (acres)</u> | <u>ERTS-1 Water Class (acres)</u> |
|--------------------|------------------------------|---|
| 1 | 1400 | 1335 |
| 2 | 1864 | 1688 |
| 3 | 245 | 171 |
| 4 | 32 | 20 |
| 5 | 15 | 9 |
| 6 | 151 | 116 |
| 7 | 146 | 112 |
| 8 | 40 | 28 |
| 9 | 45 | 32 |
| 10 | 102 | 80 |
| 11 | 56 | 35 |
| 12 | 48 | 37 |
| 13 | 79 | 62 |
| 14 | 104 | 80 |
| 15 | 94 | 77 |
| 16 | 19 | 11 |
| 17 | 28 | 17 |

It seems reasonable that this under-estimation should be expected because of the coarse spatial resolution of ERTS-1. It is obvious that if a resolution element of ERTS-1 partially

covers an area of water and some other cover type, it will not be classified as water. It will have a spectral response that is neither that of water or that of the other cover type.

It was our belief that the ERTS-1 resolution elements that partially cover two different cover types (such as water and forest) will have a spectral response ranging from that of water (if the resolution element covers a large percent of water) to that of forest (if most of the resolution element covers the forest cover type). However, there will be a narrow range of spectral responses corresponding to approximately 50% of water and 50% of forest that will be spectrally separable from the water and forest classes. This was believed to be the case with the "water-edge" class defined by the clustering algorithm. From Table 5.4 it becomes evident that the water-edge class has a spectral response that is somewhat between that of the lake water and that of forest.

Table 5.4

Relative Spectral Response for Different Cover Types

| <u>Spectral Class</u> | ERTS-1 Bands | | | |
|-----------------------|--------------|------|------|------|
| | 4 | 5 | 6 | 7 |
| Lake Water | 25.1 | 18.3 | 12.2 | 3.3 |
| Water-edge and river | 27.3 | 20.8 | 25.6 | 12.3 |
| Agricultural | 31.0 | 24.5 | 53.8 | 32.6 |
| Forest | 29.4 | 23.3 | 42.3 | 24.4 |
| Soils | 41.9 | 45.6 | 49.5 | 21.9 |

Figure 5.7 shows two classification maps of Lake Freeman and the Tippecanoe River near Lafayette, Indiana. In one of the maps the water class (M) and the edge class (.) have been displayed, and the other shows only the water-edge class. Note



(a)



(b)

Figure 5.7 Classification map of Lake Freeman, Indiana (a) where the "water" class (M) and the "edge" class (.) are displayed. In (b) only the "edge" class has been displayed.

that some fields in the area where flood water was present were classified as water-edge. Close inspection of the available low altitude photography showed that those fields classified as water-edge were areas inundated by recent precipitation. It should be noted that the narrow Tippecanoe River downstream from the reservoir was also classified as water-edge. This river is approximately 70 meters wide at the point in the classification map where two resolution elements delineate the river course.

Figure 5.8 shows four possible cases in which the resolution element of the sensing system covers different percentages of the lake surface and adjacent forest. Because the spatial resolution of the ERTS-1 scanner is such that each resolution element covers an area of approximately 184' x 256', the spectral response from the resolution element that partially covers two different cover types will be an integrated (average) value of that of each of the two cover types. In essence, this is a simplified two-class case of the more general problem of "classification of unresolved objects" which has been extensively considered from a theoretical point of view. If the proportion of Class " c_i " in the resolution cell is " p_i " and its mean and covariance matrix are " μ_i " and " Q_i " respectively, since the pure signatures of the individual classes are taken to be Gaussian, it can be shown that the distribution associated with the " p " is also Gaussian.* Thus, the statistics of the spectral class combination will be

$$\mu_p = \sum_i p_i \mu_i$$

$$Q_p = \sum_i p_i Q_i$$

* "Classification of Unresolved Objects", Technical Report by TELESPAZIO, ESRO CR-297, Rome, Italy, 1973.

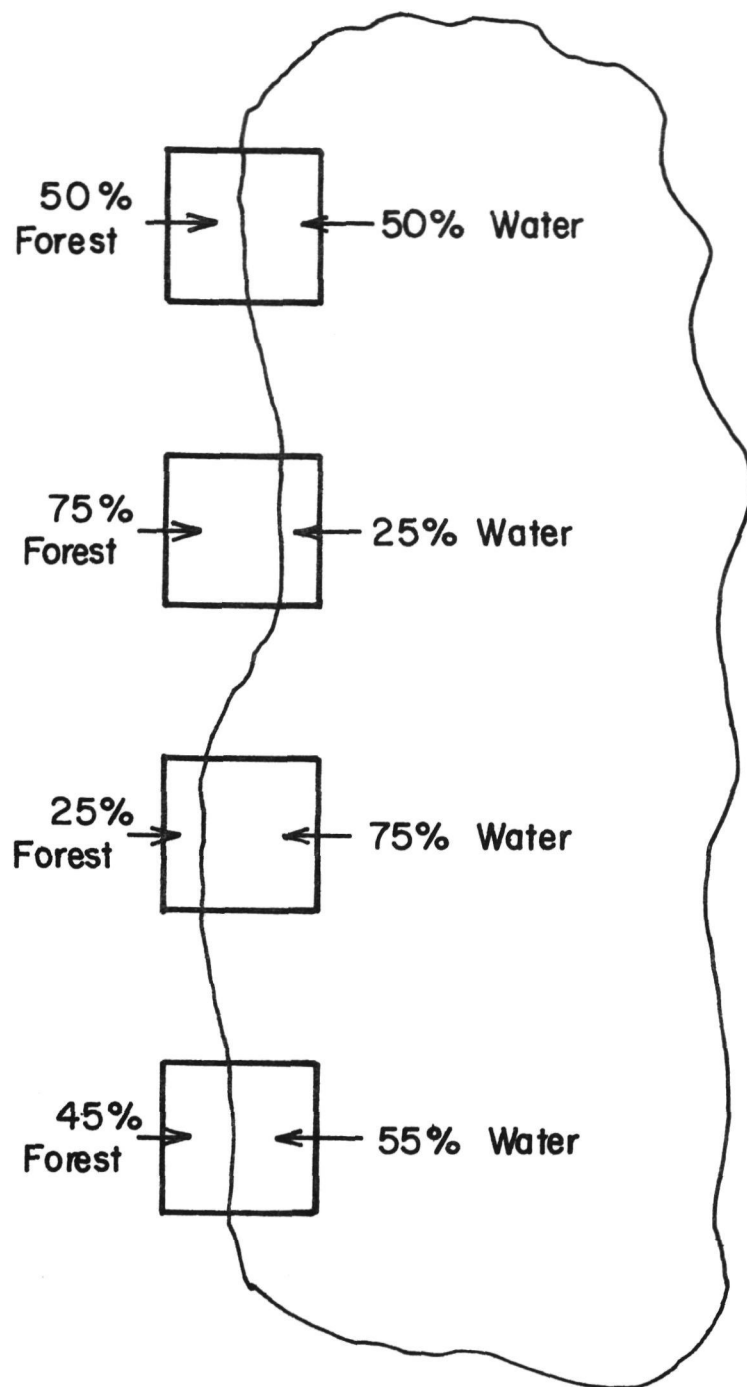


Figure 5.8. Hypothetical lake surrounded by forest. The squares represent hypothetical resolution elements.

targets covered by the resolution element. Table 5.5 shows the results of the weighted average spectral response for the four possible cases illustrated in Figure 5.8.

Table 5.5

| <u>Weighted Average Spectral Response</u> | | <u>ERTS-1 Bands</u> | | | |
|---|-----|---------------------|----------|----------|----------|
| | | <u>4</u> | <u>5</u> | <u>6</u> | <u>7</u> |
| <u>Combinations of Forest and Water</u> | | | | | |
| 50% | 50% | 27.3 | 20.8 | 27.3 | 13.9 |
| 75% | 25% | 28.3 | 22.1 | 35.2 | 19.1 |
| 25% | 75% | 26.2 | 19.6 | 19.7 | 8.1 |
| 45% | 55% | 27.1 | 21.0 | 25.2 | 12.8 |

Comparison of the weighted average spectral responses shown in Table 5.5 with that of the water-edge class defined by the clustering algorithm indicates that a resolution element that contains approximately 45% forest and 55% water would produce a spectral response similar to that of the "edge" class. This comparison is illustrated in Table 5.6.

Table 5.6

Comparison of the "Edge" and Weighted Average Spectral Responses for Resolution Elements Covering Approximately 45% Forest and 55% Water

| | <u>ERTS-1 Bands</u> | | | |
|---|---------------------|----------|----------|----------|
| | <u>4</u> | <u>5</u> | <u>6</u> | <u>7</u> |
| Actual Water-Edge Class (Lake Freeman Data) | 27.3 | 20.8 | 25.6 | 12.3 |
| Calculated Water-Edge Class (Weighted Average, Using 45% Forest, 55% Water) | 27.1 | 21.0 | 25.2 | 12.8 |

From the above considerations, it follows that to improve the water acreage estimates from ERTS-1 data, one should apply a correction factor that would account for the water surface area that is not classified as water, but as "water-edge". These considerations also imply that the fraction of the total edge class to be added to the water class would be approximately fifty percent. In order to test the significance and validity of this correction, a statistical analysis was performed.

The first step in the statistical analysis was to calculate the correlation coefficients between the estimated acreages and the USGS figures. The results showed that the estimated acreages from ERTS-1 data were highly correlated (correlation coefficients ≈ 0.99) to the standard USGS values, regardless if one counted the water class only or if 50% of the edge class was added to the water class. However, this statistical analysis did not provide us with an indication of how close the uncorrected and corrected estimations were to the USGS standard acreage values. Thus, an analysis of variance (ANOVA) was conducted, and an F-test was performed to show whether differences among several means (in our case, between the USGS method and every other method) are significant. The two different methods of acreage estimation that were compared with the USGS figures were,

- Method 1 - counting the water class only
- Method 2 - counting the water class plus 50%
of the edge class.

The actual test utilized in our analysis was the LSD (Least Significant Difference) test.

Declare \bar{y}_0 and \bar{y}_j significantly different at a level α if:

$$|\bar{y}_0 - \bar{y}_j| > t_\alpha \sqrt{2} \quad S\bar{y}$$

where,

\bar{y}_0 = USGS Method

\bar{y}_j = all other methods ($j = 1, 2$)

t_α = value from the student-t distribution for a level of significance $\alpha = 0.01$ (two tailed test) and with the same degrees of freedom on which S^2 is based.

and $S\bar{y}$ is defined as follows:

$$S\bar{y} = \sqrt{\frac{S^2}{n}}$$

where,

S^2 = MSE from ANOVA

n = number of samples on which the means are based.

Thus, the LSD test value was computed. The test value obtained for the comparison of the two methods of estimation with the USGS figures was,

$$t_\alpha \sqrt{2} \quad S\bar{y} = 27.8018$$

and the values of $|\bar{y}_0 - \bar{y}_j|$ for methods 1 and 2 respectively are shown in Table 5.7.

Table 5.7
Results of LSD Test for the Two Methods of
Acreage Estimation from ERTS-1 Data

$$|y_0 - y_1| = 30.3530$$

Test value = 27.8018

$$|y_0 - y_2| = 7.9412$$

Inspection of the results shown in Table 5.7 indicate that the Least Significant Difference between the USGS figures and the estimated acres using ERTS-1 data correspond to method 2. That is when, in addition to the water class, one considers one-half the number of edge-class acres and counts them as water. These results are consistent with the previously reported results that the spectral response of the water-edge class must consist of a combination of the spectral response of approximately one-half an area on the ground covered by water (per resolution element). Furthermore, the results of this test indicated that even though there is a high correlation between the standard USGS figures and those obtained by counting the water class only, there is a statistically significant difference between the acreage figures at the .99 confidence level.

Although consideration of the edge class for the correction of acreage estimates from ERTS-1 data yielded accurate results, it is clear that the definition of the water-edge class requires sophisticated processing techniques, such as a clustering processor. Although this appears to be a satisfactory approach to accurately determining the water area, it was felt that a less complex approach to developing a correction function could be developed. Thus, the following procedure was followed.

The data in Table 5.3 was plotted on linear graph paper with the actual (USGS) sizes along the ordinate (vertical axis) and the estimated sizes (from the ERTS-1 data) plotted along the abscissa (horizontal axis). The data in Table 5.3 was divided into two categories: (1) lakes smaller than 100 acres and (2) lakes larger than 100 acres. The resulting graphs are shown in Figures 5.9 and 5.10 from which one can determine the actual size of a lake for which one knows the estimated size from the ERTS-1 data.

It is also interesting to note that the percent under-estimation (error) from the satellite data decreases as the size of the lake increases. This occurs because the border pixels (edge

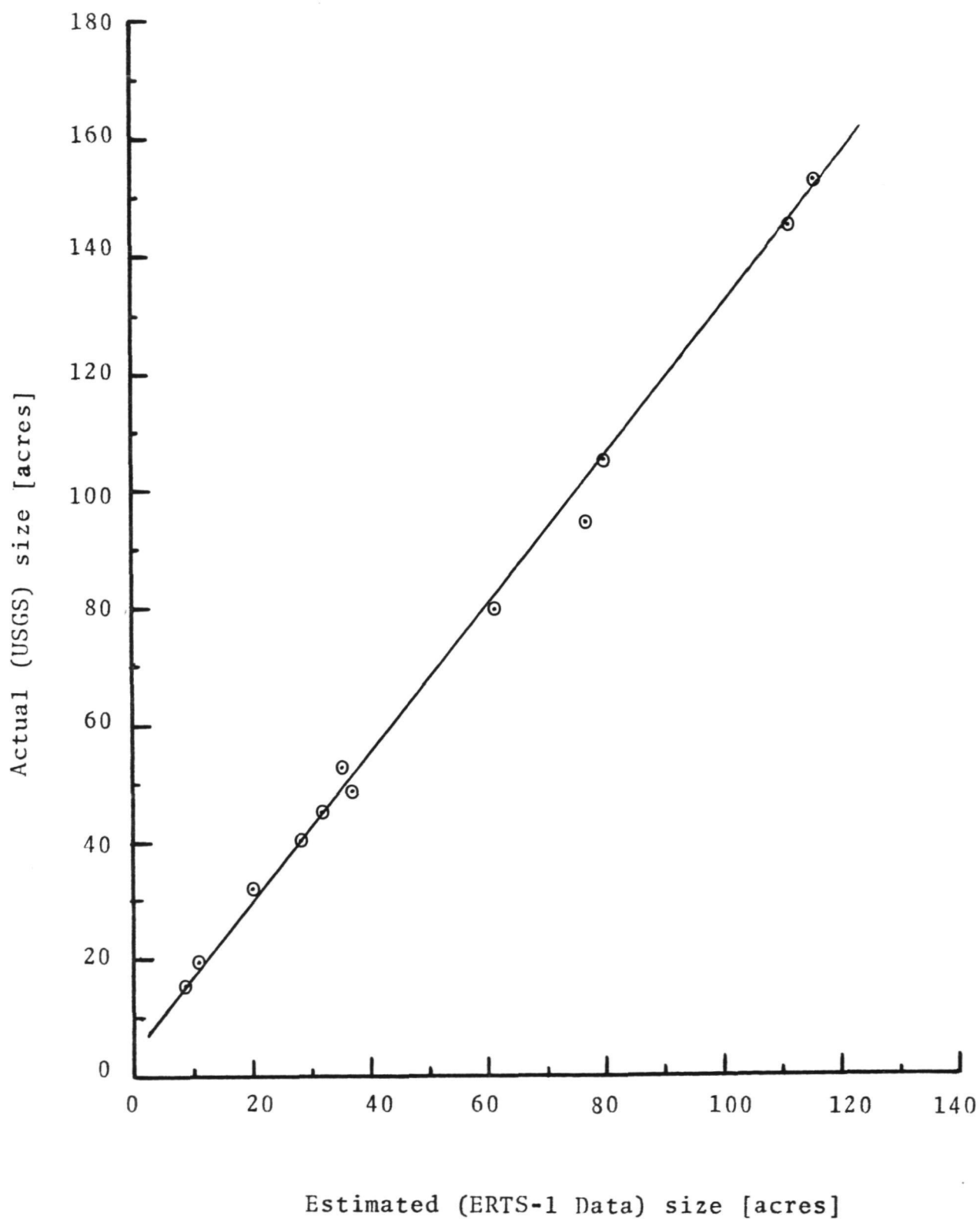


Figure 5.9. Estimated acreages (from ERTS-1 data) versus actual (USGS) acreages. For lakes less than 100 acres in size.

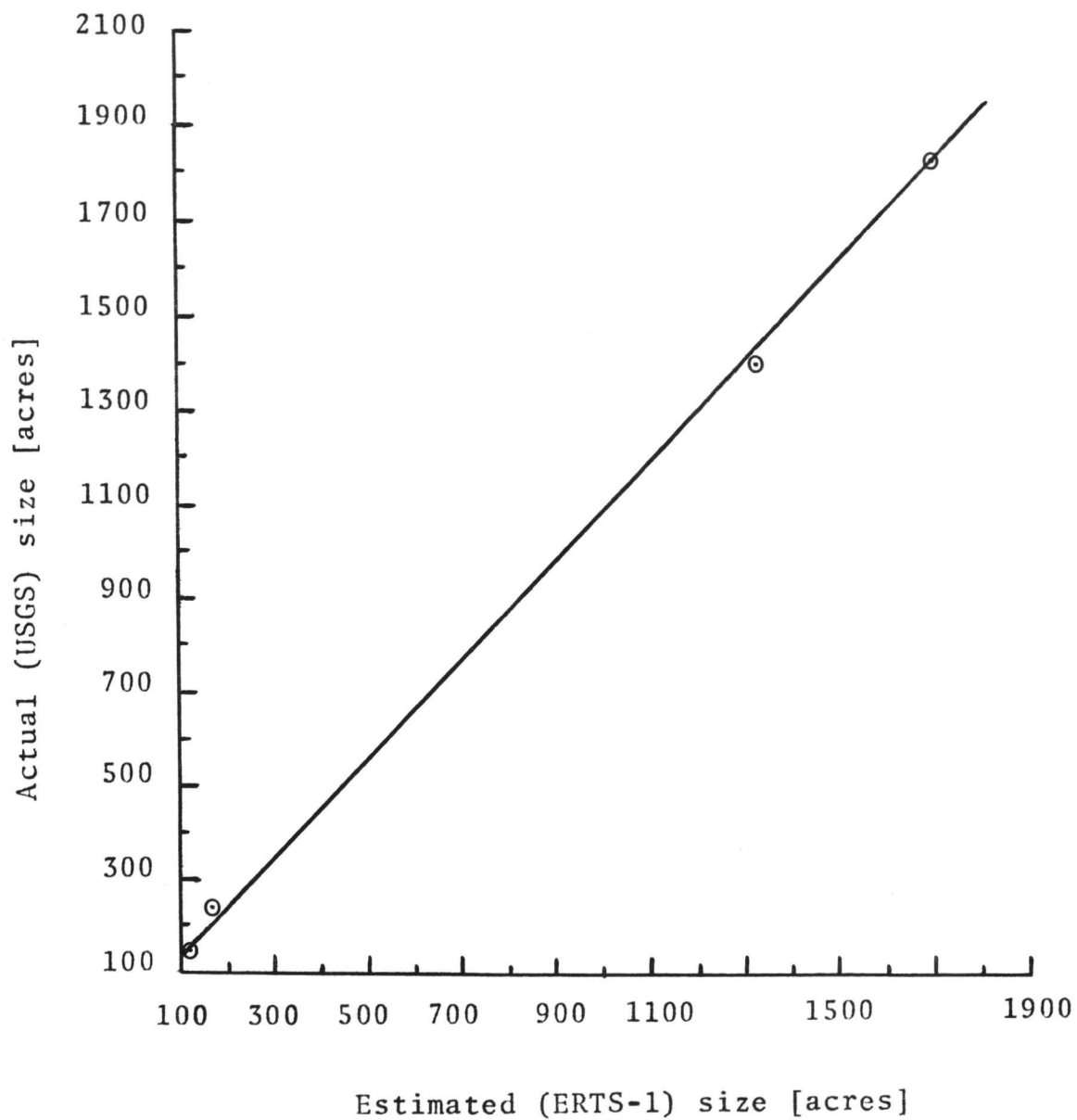


Figure 5.10. Estimated acreages (from ERTS-1 data) versus actual (USGS) acreages. For lakes larger than 100 acres in size.

resolution elements) constitute a smaller percent of the total area for larger lakes. This trend is illustrated in Table 5.8.

Table 5.8

Estimated Acreages from ERTS-1 Data and
Percent Under-estimation

| <u>Lake</u> | <u>Actual size (acres)</u> | <u>Estimated size (acres)</u> | <u>Under-estimation (%)</u> |
|--------------|--------------------------------|-----------------------------------|-----------------------------|
| Fish | 15 | 9 | 66 |
| Muskelonge | 32 | 20 | 60 |
| Caldwell | 45 | 32 | 41 |
| Yellow Creek | 151 | 116 | 30 |
| Bass | 1400 | 1335 | 5 |

5.7 Conclusions and Recommendations

The results reported in the previous sections indicate that ERTS-1 multispectral data and computer-aided classification techniques can be utilized to detect and map different spectral classes of surface-water which may correspond to different levels of turbidity. However, it is clear that more work in conjunction with collection of more accurate and reliable field observations are needed in this area of research. Nevertheless, from previously reported work on turbidity (Weisblatt, et al, 1973) with ERTS-1 and EXOTECH field spectroradiometer data and from existing processing and analysis techniques at LARS, such as the "Layered Classifier", it seems feasible to be able to map and make quantitative determinations of water turbidity levels. The procedure recommended for the quantitative determination of the amount of suspended solids present in lakes and reservoirs is to utilize a layered classification scheme in which the first step would be to use all four spectral bands of ERTS for the separation of water from every other cover type through a maximum likelihood classifi-

cation. Then, the second step would consist of a level slicing technique applied to only one spectral band, such as band 5 (0.6 - 0.7 μ m) which has been shown to have a linear response as a function of turbidity levels.

In the case of the aircraft data analysis, the results indicate that under certain conditions the sun-scanner-look angle effect is so pronounced that the data is useless for surface-water studies because any spectral characteristic due to either depth, turbidity or any other water quality parameter would be completely masked by the strong specular reflections from the water surface. Therefore, careful aircraft mission planning is needed in order to avoid the sun-scanner-look angle effects.

Finally, it was shown that because of the coarse spatial resolution of the ERTS-1 sensor system, there is a consistent under-estimation of the surface area of water bodies. However, two methods to correct the estimated sizes of lakes were developed. One considers the "water-edge" spectral class, and the other consists of a correction graph (Figures 5.9 and 5.10). The resulting water acreage estimations using ERTS-1 MSS data together with computer-aided processing techniques, have been shown to be statistically correlated to the standard USGS data. Thus, one may conclude that it is possible to accurately estimate the size of lakes and reservoirs from ERTS-1 data, provided an appropriate correction factor is applied.

6. Earth Surface Features Identification

6.1 Introduction

Research on the applicability of ERTS-1 data to the regional land use planning process suggests that certain data needs of planning groups, that of specific earth surface features must be effectively realized. During the period of this research project a study was conducted to determine if ERTS-1 data could be used to supply this necessary information.

6.2 Background

Concern for environmental resource planning, as well as the demands of society asking that physical planners predict environmental consequences in quantified terms before development, dictates the need for better data in the land use planning process. This need for relevant and environmentally responsive regional planning and management is crucial in a state's drive for maintaining a high level quality of life. An understanding of our environment as a complex interactive entity has historically been neglected, disregarded or not understood by land use decision makers. Besides the need for a better perception of the environment, intelligent land use decision making is extremely difficult at the present due to the static or nonexistent and archaic information resulting from data inventory techniques that are slow and antiquated. The decision maker/planner typically lacks relatable basic information about the use, the composition, character and temporal dynamic qualities of landscape change. What is also not understood is that this data is not impossible to obtain. Some of the most basic forms of data such as the extent and variation of vegetation cover, wetland distribution, urban growth patterns, as well as the general character of the landscape are examples of data that have traditionally not been available to the regional decision maker.

In order that this decision making process be optimized, it is imperative that these landscape elements be identified, analyzed and understood. As a means of accomplishing this end, research was conducted into the utilization of remote sensing techniques as a means of acquiring this data as well as relating its influence to land use planning. The process of environmental decision making becomes critical to controlling the tremendous amount of environmental impact upon our landscape due to urban and social population pressures.

The future holds that these numerous land uses can place high demands on our physical environment thus necessitating means of monitoring existing and unique data so that future management decisions can be made. Management and land use decision making becomes imperative in establishing and maintaining a compatible relationship of the diversity of elements in our environment. It therefore becomes necessary - in order to make accurate and feasible decisions relative to optimizing land use and management policies - that the decision maker have the capability of gathering and analyzing as many data variables as can be input into analysis systems and not just rely on subjective judgements.

If regional planning is ever to incorporate environmental resources into the planning process, an efficient mechanism for the inclusion of relevant, reliable data must be developed. The intent of this research investigation was thus to evaluate and document the hopefully potential applications of ERTS imagery to this need. It must be realized that ERTS data will not encompass all the regional planning needs but it does offer a technique by which the data acquisition process can be improved.

6.3 Goals and Objectives

The basic procedure was to investigate the potential of ERTS-1 data for regional land use planning with the specific purpose of identifying those natural and cultural earth surface features significant to the decision making process. Imagery analysis will include the spatial accuracy and quantification of critical resources.

Initial objectives are to:

- A. Compare those earth surface features identified on ERTS-1 imagery to specific natural data previously determined. (Considering scale and change over time.)
- B. Determine the usefulness of this data to the land use decision making process (as determined by the user).
- C. Utilize computer capabilities in delineating and classifying specific earth surface features.
- D. To generate computer maps displaying earth surface features identified.
- E. Spatial quantification of delineated data.

6.4 Approach

In order to accomplish the basic objectives indicated, the data variables critical to decision making must be identified and classified. It was therefore essential that the prime users were identified and their basic needs be categorized. The initial investigation consisted of basic interpretation of those natural and cultural earth surface features identifiable and critical to land use planning. Comparisons were made with an existing data base to establish the validity of the imagery in extracting these basic resources.

6.41 Cooperating Agencies

The complexities of such a problem as ERTS data analysis requires the corporate efforts of many individuals and agencies. Therefore, it was essential that work done within the earth surface feature identification program be coordinated with ongoing projects in LARS laboratory as well as those agencies considered as prime users. At present the agency which appears to have most

interest in this ERTS project is the Indiana Department of Natural Resources. They have conducted a comprehensive state-wide inventory of data that can be utilized as a data base for comparison purposes. Other cooperating agencies could be the Division of Planning and specific county planning agencies.

6.42 Study Area

The study area for this project was located in Tippecanoe County, Indiana, East of Lafayette, Indiana. (See Figure 6.1) The area encompasses a section of the county which is representative of the rural Indiana landscape and is fourteen kilometers East to West by ten kilometers North to South for a total of 140 square kilometers or 63 square miles - approximately 35,000 acres.

The data inventory process was set up to not only offer the user the opportunity to familiarize himself with techniques of regional inventory and analysis but to be utilized as a tool for further investigation into the regional management and evaluation process.

6.43 Data Identification and Storage

The selection of data variables for the data base to be utilized for comparative purposes was directed to the goal of obtaining the necessary information with which anticipated land use decisions could be made. Resource variables included the following:

I. Natural Resources

A. Hydrologic Systems

1. Streams
2. Rivers
3. Ponds
4. Drainage ways

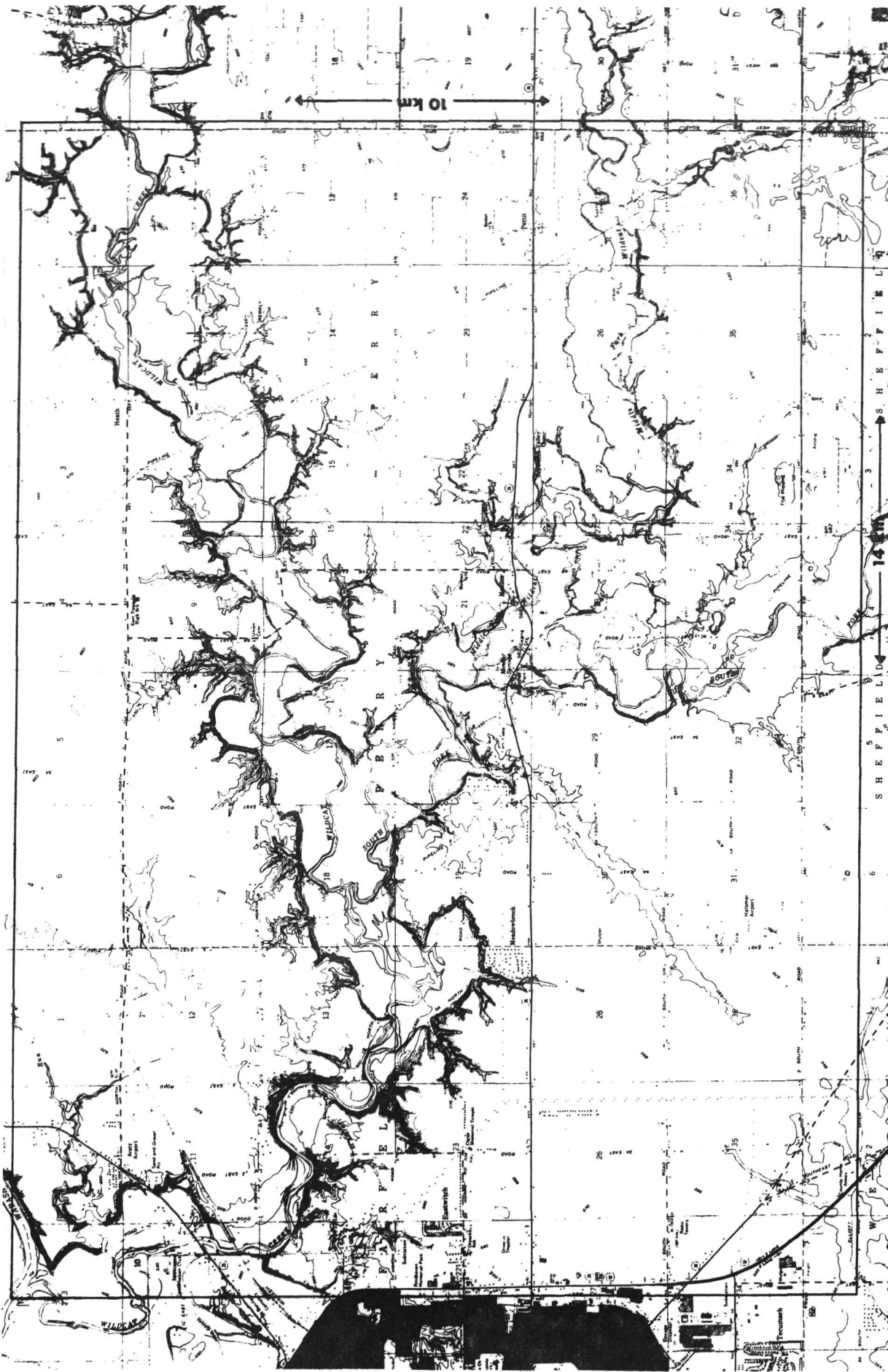


Figure 6.1 EAST CENTRAL TIPPECANOE COUNTY

B. Ecological Systems

1. Vegetation-Forest cover type
2. Lowland Forests
3. Upland Forests
4. Wetlands

C. Physiographic Systems

1. Topographic orientation
2. Topographic slope
3. Topographic elevations
4. Landforms

D. Pedological Systems

1. Soil Types
2. Erosion class
3. Subsoil characteristics
4. Bedrock characteristics
5. Flooding potential

E. Natural Landscape Units

1. Sub Watersheds
2. Ground water conditions

II. Cultural Resources

A. Existing land use systems

1. Agricultural activity
2. Residential activity
3. Development activity
4. Recreational activity

B. Communications Systems

1. Transportation type
2. Air and Rail activity
3. Utility types

C. Cultural Landscape Units

1. Property ownership
2. Zoning

These variables were sub-divided into representative component parts and extracted such as to relate them spatially within the study area. This extraction process was intended to organize and translate cartographic information at various scales into computer compatible format. The Universal Transverse Mercator (UTM) reference system was used as the framework for data extraction with each cell being one tenth of a kilometer square. Two methods were used to extract the data. First, those variables that comprised line or points, were simply recorded as to their presence or absence within a cell. The second method was to record the predominant area within the cell.

The intent of storing the data in this manner was to be able to have a means of comparing the extent of accuracy to which certain natural resource variables could be identified from ERTS-1 imagery.

6.5 Procedure

Two areas of investigation were studied to achieve the ultimate goal of attempting to establish a technique whereby ERTS data could be analyzed so as to determine its usefulness in extracting resource information critical to maximizing land use decisions.

In the first area of study a significant amount of effort was devoted to the classification process of identifying numerous natural and cultural resources from ERTS data.

A detailed land use classification consisting of 29 spectral classes of the study area was completed using LARSYS techniques. This classification indicated that a number of urban features can

be differentiated including commercial development (shopping centers), two distinct classes of new residential area, and one class of older residential area. This data was collected on September 30, 1972. Some misclassification occurred between old residential areas and row crops.

It was also determined that forest cover can be reliably differentiated from other cover types. There were three different categories of forest cover differentiated (these classes are described in more detail in a latter section of this report) -- a factor that led to the utilization of this resource variable in the second area of this investigation. The other major land-use categories which have been defined included the following: three classes of water (lake and two river classes), pasture/grass and row crops.

A temporal overlay for the test area became available late in the period and a test analysis was performed with this data. Data from September 30, October 19, and November 6 were overlayed and a test analysis to map forest cover was performed using this data set. Very little difference was found between the forest cover map from the overlayed data and that from the map using September 30 data only.

To accomplish Area Two of this investigation, it was determined to focus initial attempts on one specific resource variable -- that of forest cover as classified on ERTS data September 30, 1972. Information was first gathered and stored in a spatial data bank so as to be utilized for ground truth purposes.

Since the primary goal of this area of study was that of establishing a correlation of ERTS data to that of an existing spatial data system ERTS imagery was selected, classified, and then reformatted so as to be compatible with the existing data

analysis system. Of interest here were the forest cover groups assembled as Class T, E, and Y. These classes were manually extracted onto a base grid in an attempt to make them spatially compatible with the existing analysis system. The codes for extraction as well as the class groups are listed below:

| <u>Extraction Code:</u> | <u>ERTS Data Symbol:</u> | <u>Classification Groupings:</u> |
|-----------------------------|--|---|
| 0 | (symbolizes the lack of the three class groupings symbol) | |
| 1 | T | Trees 1 Green Agricultural Trees 5 Trees 6 |
| 2 | E | Trees 3 |
| 3 | Y | Trees 4 Dense Forest |

The existing data bank has nine levels of forest classification (based on percent density of specific forest cover) which for purposes of this initial investigation was too detailed. These nine sub-groups were thus agglomerated into three groups as follows:

| <u>Original Extraction Code</u> | <u>Agglomerated Extraction Code</u> | <u>Map Symbol and Meaning</u> |
|---|---|---------------------------------------|
| 0 | 0 | Blank (none) |
| 1 | 0 | Blank (none) |
| 2 | 1 | |
| 3 | 1 | L (lowland) |
| 4 | 1 | |
| 5 | 1 | |
| 6 | 2 | |
| 7 | 2 | U (upland forest) |
| 8 | 2 | |
| 9 | 2 | |

Figures 6.2 and 6.3 illustrates the distribution of forest cover, extracted from low, altitude data, in the study area.

The data extracted from ERTS-1 imagery was then spatially related to the data bank. Through basic logic statements and utilizing the manipulative and analytical capabilities of the data bank, comparative investigations were made studying the accuracy levels of forest cover identification vs. ERTS-1 data.

6.6 Results

Initial classification was found to produce a fairly good overlay with the topographic map and showed excellent correlation between the water areas, forest areas, agriculture and urban and commercial areas in the portion of the study area. The results have been evaluated by comparison of the classification to the information available in the data bank. The areas classified as forest from the ERTS data were then extracted from the computer printout and is presented in Figure 6.4. The two extractions were overlayed and a comparison illustration was made, Figure 6.5, showing areas of disagreement as a dark shade. Table 6.1 gives the percentage of the area classified and the various classes for four separate classifications using the same data set. The first classification used all eight data channels and produced the best overlay with the topographic map. Two additional classifications were made using the same statistics and the same data set. However in this case the channels representing September data were classified in a second classification. An additional classification was made using the "best four" channels as selected by use of the LARSYS SEPARABILITY processor. The "best four" set of channels was selected from examination of output from this processor and it was found that the majority of the top-rated 10 to 15 channels by the "SEPARABILITY" processor consisted of an infrared and a visible channel selected from each

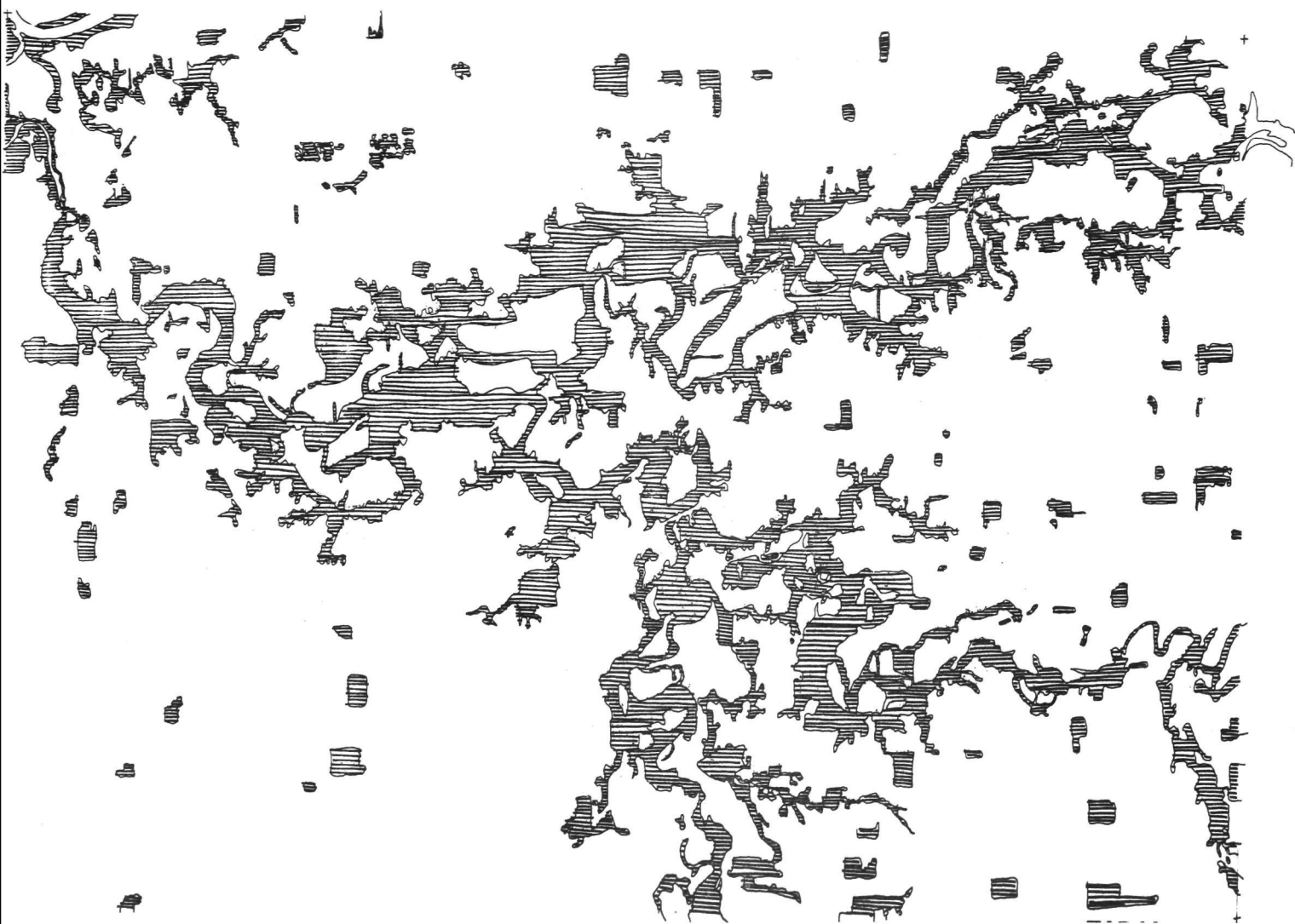


Figure 6.3 Forest cover in the test site extracted from 1971 low altitude aerial photography.

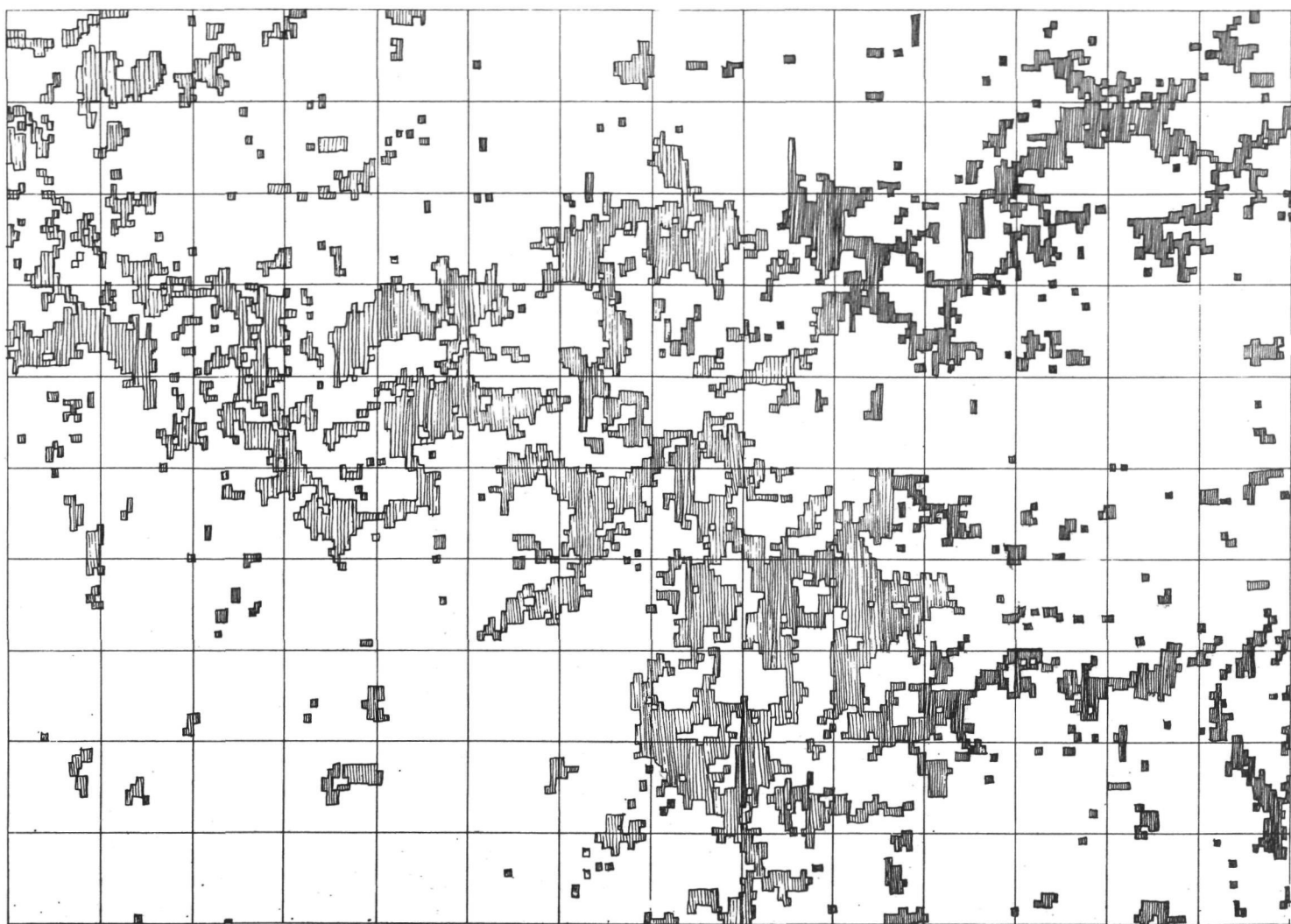


Figure 6.4 Forest cover extracted from ERTS data computer classification printout.

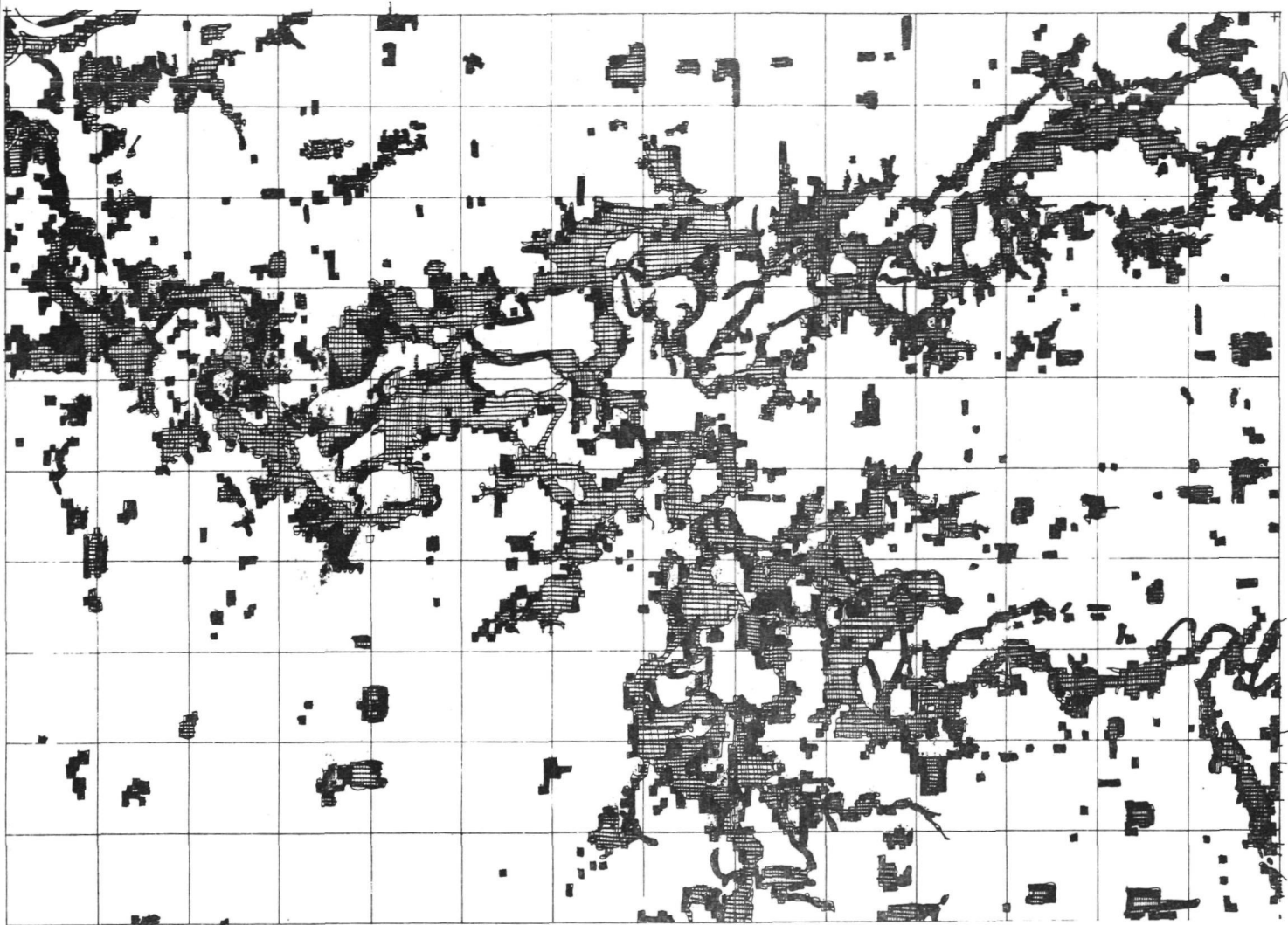


Figure 6.5 Comparison of Figure 6.3 and 6.4 showing areas of discrepancy as a dark shade and areas of forest agreement as a cross-hatched pattern.

Table 6.1. Percentage of Area Classification Using Complete Eight Channel Overlay Data Set and Subsets of Data

| <u>Data Set</u> | <u>Agriculture</u> | <u>Forest</u> | <u>Urban</u> | <u>Water</u> |
|-----------------|--------------------|---------------|--------------|--------------|
| Sept. & June | 75.7 | 18.9 | 4.9 | 0.5 |
| Sept. | 71.0 | 23.3 | 5.2 | 0.5 |
| June | 70.8 | 20.4 | 8.5 | 0.3 |
| "Best 4" | 75.8 | 18.5 | 5.3 | 0.4 |

Table 6.2. Percentage of Area Classification Using Only September Data for Entire Analysis Procedure (no temporal information in training set selection)

| <u>Agriculture</u> | <u>Forest</u> | <u>Urban</u> | <u>Water</u> |
|--------------------|---------------|--------------|--------------|
| 62.8 | 22.6 | 11.5 | 3.1 |

date. The relative evaluation of these channels indicated that they were essentially equivalent in their ability to separate the classes existing in the training sets.

The interesting point of this table is that while some loss in accuracy in agriculture and urban and forest classes existed in September data alone and also in the June data alone, the results for the best four channels are essentially identical to those of the complete eight channel data set. This suggests the possibility that temporal overlays might be used on small areas to produce training sets which could then be used for classification using only four channels is encouraging since the computer time required for four channels is far less than that required for eight.

Table 6.2 shows the results of separate classification made using only September data but in this case the entire classification process used only September data, therefore excluding any temporal information from the definition of training sets. The same geographic area was classified and evaluation was made from aerial photography in the same manner as the previous classification. However, the classification contains a different number of classes and the results are shown to be quite different from any of the previous four. This classification exhibits the problems which have been encountered with many classifications in which resource information have been mixed together.

The comparison of the two tables indicated the value of temporal information in both definition of training sets and in classification.

In the first comparisons of ERTS-1 data to that of low altitude data it was seen that the number of acres identified as forest on ERTS was more than that on the low altitude data.

The comparison generated an initial accuracy of 70 - 75%. This relatively low result was attributed in part to the initial inaccuracies in classification and resolution capabilities of ERTS.

With this in mind more sophisticated classification of an ERTS image was produced and the ground truth comparison data re-extracted. Upon evaluation of this data the accuracy was increased to beyond 85%. Thus the resource identification of forest cover in the study area appeared to be relatively accurate but for the information to be utilized for detailed land use inventory and analysis this same accuracy must be achieved in spatial location of the information.

A comparison was thus made of the spatial accuracy of the information comparing two scales of data entry - a one tenth kilometer square and a one fifth kilometer square. Figure 6.6 represents existing forest cover in the study area as extracted on one tenth kilometer cell from low altitude aerial photography. This data was compared to Figure 6.7, which is forest cover as extracted from classified ERTS data on a one tenth kilometer cell, and is represented in Figure 6.8. The dark areas of Figure 6.8 represent the regions of discrepancy which, in the most part, occur at the fringe areas. The spatial accuracy achieved at this scale of data entry was approximately 84%, which was less than the overall accuracy. There were 11,859 corresponding cells and only 2,141 cells in conflict. Utilizing this same approach but with the data entry scale at one fifth kilometer (Figures 6.9, 6.10, and 6.11) there were 3,119 corresponding cells and 381 cells in conflict for a spatial accuracy of approximately 90%.

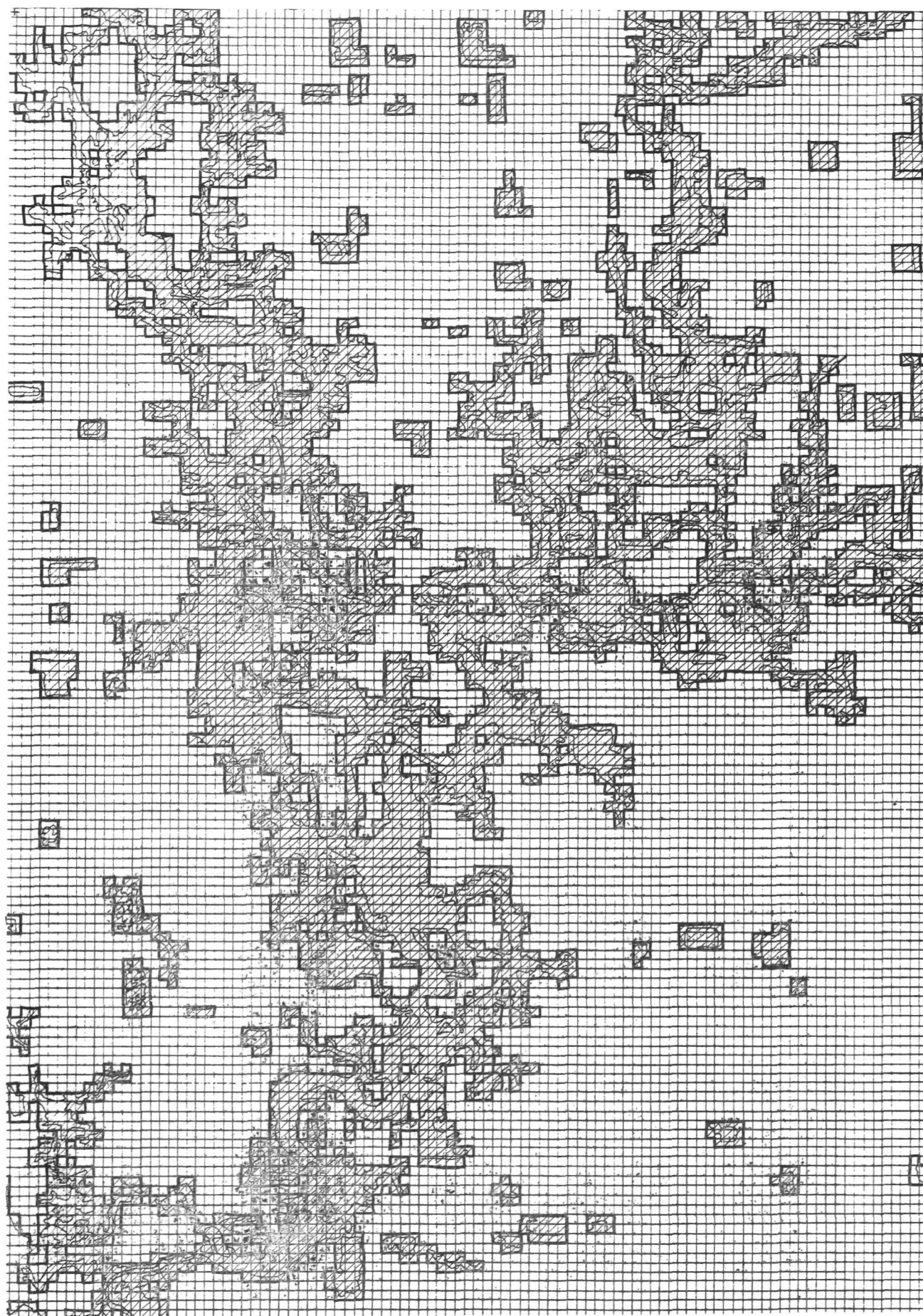


Figure 6.6 Existing Forest Cover-Extracted at One Tenth Kilometer Square.

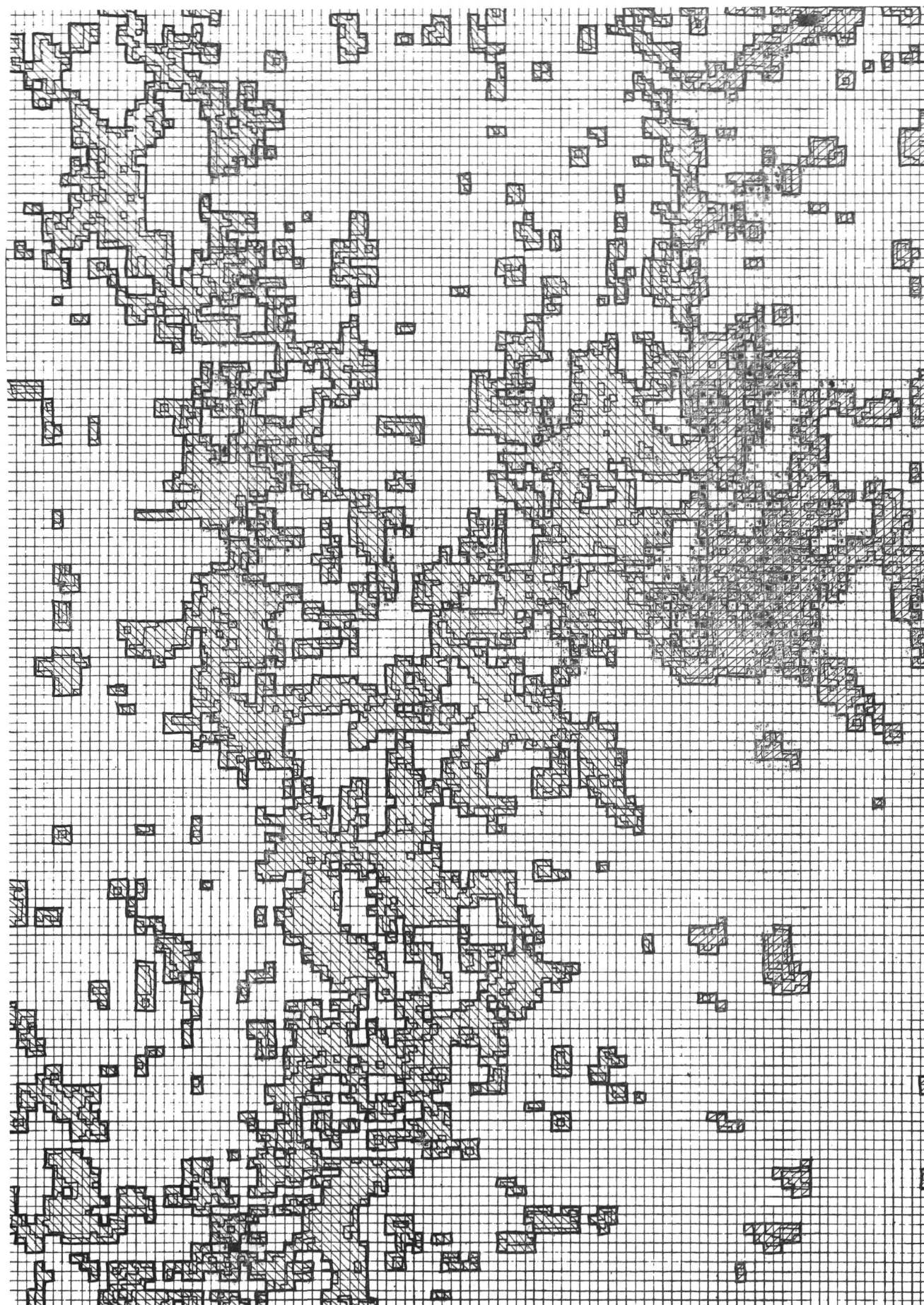


Figure 6.7 Forest Cover Extracted From ERTS at One Tenth Kilometer Square



Figure 6.8 Comparison Between Existing and ERTS Data at One Tenth Kilometer Square. Dark Areas Represent Areas of Non-Correlation. 84% Correlation Accuracy.

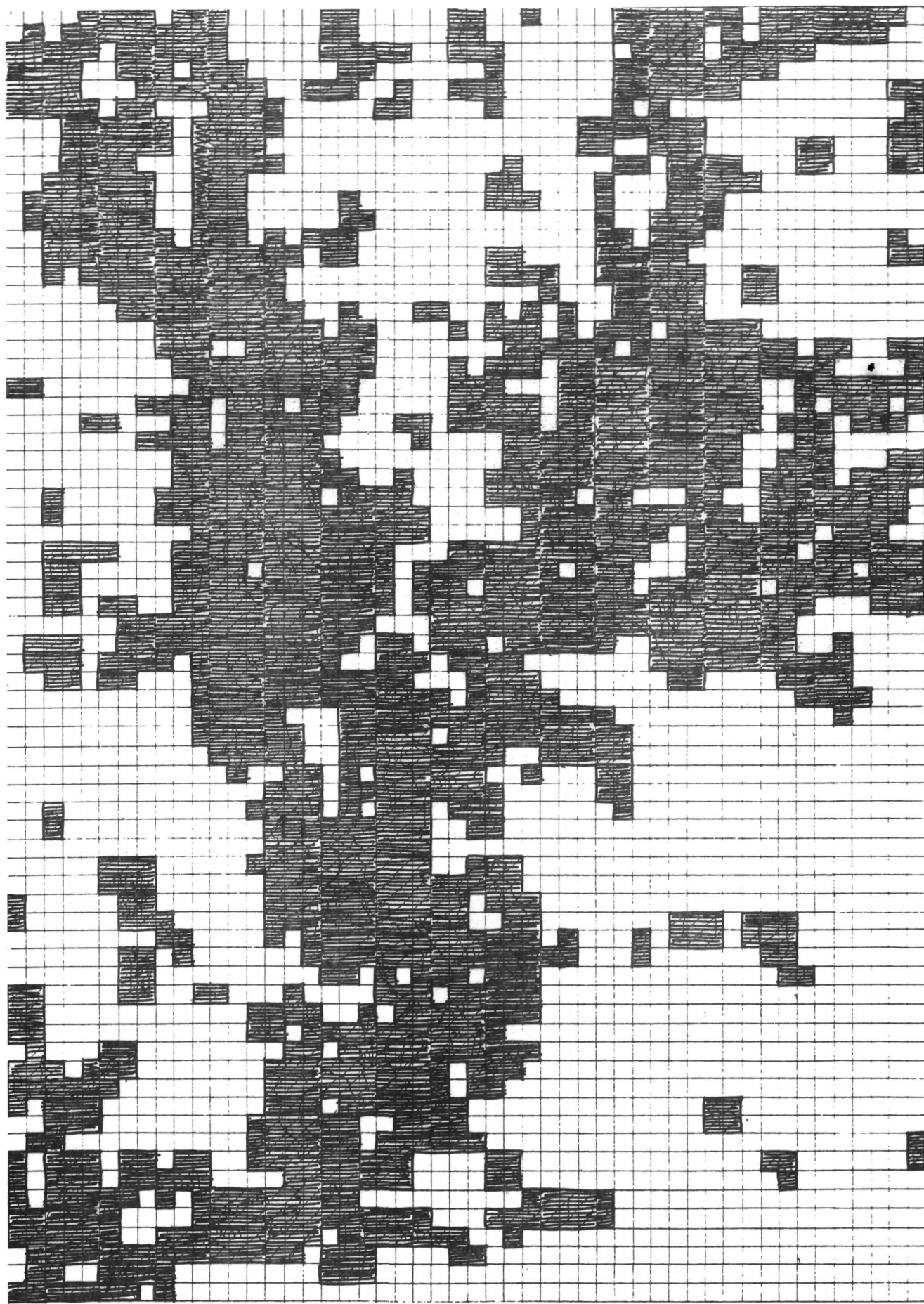


Figure 6.9 Existing Forest Cover-Extracted at One Fifth Kilometer Square.

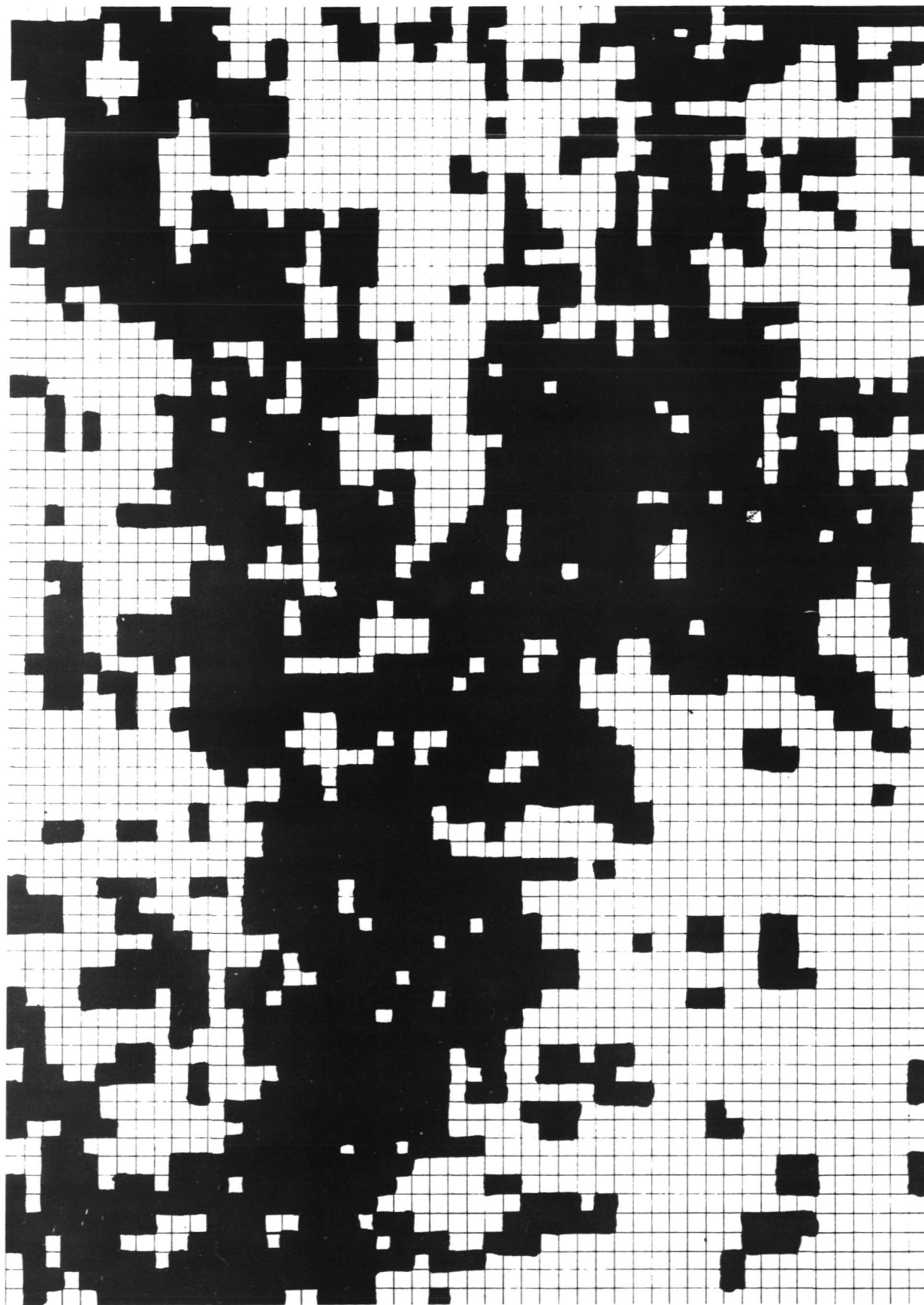


Figure 6.10 Forest Cover Extracted from ERTS at One Fifth
Kilometer Square.



Figure 6.11 Comparison Between Existing and ERTS Data at One Fifth Kilometer Square. Dark Areas Represent Areas of Non-Correlation. 90% Correlation Accuracy.

6.7 Conclusions

It was thus concluded that by increasing the size of data entry cell greater spatial accuracy could be achieved with an overall identification accuracy of 85%. It is also presumed that the overall identification accuracy will also become greater with more investigation into temporal overlay classification which will begin to make ERTS a valuable source of gathering earth surface features for regional resource planning.

Attempts were also made at automatically inserting ERTS data into a data bank but due to initial registration problems the spatial accuracy was greatly reduced.

Although this investigation was limited to comparison of only forest cover data, other earth surface resources could be identified and analyzed by the same process. The spatial accuracy results are encouraging and the future holds that with the achievement of a more accurate automatic ERTS data entry a semi-automatic system of data bank development for utilization in the land use planning process can result.

7.0 Analysis Technique Development

7.1 Introduction

The novel character of the ERTS data, particularly the quantity, resolution and large area coverage, anticipated even before launch, suggested some areas in which advancement of the data analysis technology could have particularly significant impact. These included the development of data-based criteria for defining ground cover classes and selecting training samples; development of adaptive pattern recognition techniques which would make use of scene context, particularly in a spatially changing environment; and implementation of pattern recognition algorithms using layered decision logic to make the analysis process more efficient. Progress has been made and new knowledge uncovered in all three areas as the result of research under this contract.

7.2 Data-Based Criteria for Defining Training Classes

The scale and resolution of the digital ERTS-1 MSS data have, as expected, led to multispectral data analysis techniques which differ significantly in detail (if not in concept) from the techniques developed for data collection by aircraft. Interdisciplinary efforts at LARS involving both data processing and applications scientists have evolved techniques applicable to a wide range of earth survey problems (most of which are evidenced in this report).

Two extreme situations may be identified which require somewhat different analysis approaches. In one case no ground observation data are available -- or at best ground observation data of a very general nature (such as might be obtained from interpretation of high altitude underflight photography). At the other extreme is the case in which detailed ground observation data are available (for instance, from field visitation). The analysis techniques for these cases differ primarily in the

extent to which training class definition depends on spectral variability inherent in the multispectral data. Following are outlines of the analysis procedures applicable to the two extreme situations.

Case I: Limited ground observation data.

1. Apply cluster analysis to randomly selected areas and/or areas known or expected to contain cover types of interests.
2. Associate the clusters ("spectral classes") with general ground cover types. Interpret relative response in one or more channels to indicate general cover type.
3. Classify the image using the spectral class definitions obtained in the previous step.
4. Refine the spectral class associations if possible based on any available information about the scene.
5. Perform qualitative evaluation of the results.

This procedure is most frequently useful for mapping general ground cover types (e.g. bare soil, vegetation, water) as opposed to cases requiring relatively difficult discriminations (e.g. crop species with canopies having similar spectral characteristics).

Case II: Detailed ground observation data.

1. Unsupervised multichannel image enhancement
 - a. Apply cluster analysis to areas containing ground observations.
 - b. Use the spectral classes produced by the cluster analysis as a basis for classifying the scene. Use grey scale symbols to represent the spectral classes. Result: enhanced field and object boundaries. "Multichannel image enhancement" is essentially the procedure described as Case I above.

The objective of this step is to combine the spectral information from multiple channels into a single display which contains composite information from the individual channels.

2. Locate training and test samples in the enhanced imagery.
3. Use cluster analysis to refine training field selection.
4. Classify the image using the training class definitions.
5. Perform quantitative results evaluation using test samples.

The analysis approach developed for Case I is basically unsupervised classification whereas for Case II supervised classification is utilized. Of course the two methods are usually blended appropriately according to the amount and detail of the ground observations available.

A key aspect in the use of cluster analysis for the purpose of determining spectral classes is the definition of "cluster" or "class". Intuitively one feels that "cluster" is a fairly well-defined concept: given a one or two-dimensional plot of data points, one can usually discern visually any tendency of the data to be organized in clumps or clusters. Mathematically, however, it is not at all obvious how to characterize such clustering tendencies, and the situation becomes even less tractable when multivariate data are involved (often with unknown or unquantified relationships existing between the variables). Most of the research in clustering methods is involved with the formulation and justification of criteria for defining what constitutes a cluster. This is for all practical purposes an impossible task, however, unless considered in the context of the specific problem to be solved by the cluster analysis. It is one thing to use clustering to isolate multiple modes in a distribution of data; it is quite another to attempt to isolate classes of soils, say, for which spectral differences may be

indicative of significant physiochemical differences.

Because of problem dependencies of the nature cited above, the strategy has been to make a rather general cluster analysis facility available to the data analyst and to expect the analyst to perform a considerable amount of interpretation of the results based on his experience and the details of the problem. As larger and larger quantities of data are analyzed, however, the analyst needs more help in systematizing his interpretation. And ideally, of course, it would be desirable to have the interpretation as completely automated as possible.

To this end, the following cluster analysis procedure has been formulated and tested. Assume the data have been subjected to a "standard" clustering process without splitting or merging of cluster (see, for example, [1]). Since only merging will be used in the subsequent analysis (no cluster splitting), the number of clusters requested of the clustering process should be greater than the actual number needed.

Assume there are n clusters, and let d_{ij} , $i = 1, 2, \dots, n$; $j = 1, 2, \dots, n$ be the pairwise distances (Swain-Fu distances) [1] between the clusters. Let C_i be the cluster group (C-group) to which cluster i belongs.

1. Initially assign each cluster to its own cluster group, C_1, C_2, \dots, C_n .
2. Order the d_{ij} 's from smallest to largest and work through the list of d_{ij} 's from smallest to largest, as follows.

[1] Swain, P. H., "Pattern Recognition: A basis for Remote Sensing Data Analysis", LARS Information Note 111572, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, November 1972.

3. If $d_{xy} > T$, stop (merging is complete), where T is an analyst-supplied threshold value (discussed later).
4. If clusters x and y belong to the same C-group ($C_x = C_y$), go on to the next value of d_{xy} (return to step 3).
5. Compute the average distance \bar{d}_{xu} between C_x and each other C-group $C_u \neq C_x$ for which $d_{ab} \leq 0.75$ for all a in C_x and b in C_u (the average distance between C-groups is defined as the average of all pairwise distances between individual clusters in the different C-groups). Similarly compute the average distance \bar{d}_{uy} between C_y and each other C-group $C_u \neq C_y$ for which d_{ab} is ≤ 0.75 for all a in C_u and b in C_y .
 - (a) if \bar{d}_{xy} is the smallest of all of the intergroup distances so computed, then assign both C_x and C_y to the same C-group, i.e., $C_x = C_y = \text{MIN}(C_x, C_y)$. Select the next d_{xy} and return to step 3.
 - (b) otherwise, simply select the next largest d_{xy} and return to step 3.

This procedure provides a systematic means for interpreting the separability information, minimizing the total number of subclasses produced while ensuring that multimodal class distributions are avoided.

What the preceding algorithm accomplishes may be described as follows. Every cluster produced by the initial clustering phase is characterized by its mean and covariance matrix (i.e., position and dispersion). Then each cluster which is "closer than T " to another cluster or group of clusters is merged or

associated with the nearest cluster or cluster group (the definition of distance between clusters and cluster groups is contained in the algorithm description).

Once the user has specified two parameters (the initial number of clusters n , and the merging threshold T), the entire process is completely defined. Since both n and T are best determined by the nature of the problem, this allows necessary flexibility in the analysis process while providing a systematic approach which is free of analyst error and is repeatable. Selection of appropriate values for n and T is learned quickly through experience with the process. Typically n is chosen as twice the anticipated final number of clusters. For moderately difficult discrimination problems such as crop classification, $T=0.75$ is effective. Neither selection seems to be very critical.

Although the utility of the cluster merging algorithm has been demonstrated through use, a quantitative evaluation is desirable and should be pursued.

The approach to cluster analysis described above has been extensively applied in all investigations reported herein.

7.3 Adaptive Classification

Because of the large number of factors which influence the spectral characteristics of any ground cover, the characteristics may be expected to change as a function of geographical location. As a result, "retraining" of the classifier will inevitably be required when the area of coverage (i.e., the area to be classified) becomes sufficiently large. Typically, however, the spectral characteristics vary slowly, so that it may be feasible to employ adaptive techniques to automatically "track" the changes and thereby obviate the need for completely retraining the classifier. In practice what would be done is to design the classifier in such a way that it can automatically update the statistics associated with the classes. A classifier with this ability is called an adaptive classifier.

The usual approach to adaptive classification is to assume "supervised" adaptation, in which the true classification of the data to be used for adaptation must be known. But since for remote sensing applications this could require an unreasonable amount of ground observation data, it was decided in the present case to develop an adaptive model based on "unsupervised" adaptation. Such a model has been formalized and applied to MSS data with promising results. The details, which are contained in [2] are summarized in this section.

The simplest form of an unsupervised adaptive classifier would use the data associated with every classification to update the class statistics. However, it is easy to show that such an

-
- [2] Robertson, T. V. and Swain, P. H., "A Model for Adaptive Classification", LARS Information Note 050174, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette Indiana; in preparation.

adaptive classifier may be unstable in the sense that there is a high probability that it will eventually be led to classify everything, regardless of its true identity, into one class. This is true, in particular, for the familiar maximum likelihood classifier which assumes Gaussian (normal) statistics. The model developed here assumes a special case of Gaussian statistics and avoids the instability problem.

It is assumed that n -dimensional measurement vectors $X = [x_1, x_2, \dots, x_n]^T$ are to be classified into m classes $\omega_1, \omega_2, \dots, \omega_m$. The classes are assumed to be characterized by multivariate Gaussian probability density functions with unequal means and identical covariance matrices.

$$X \in \omega_i \rightarrow X \sim N(M_i, S)$$

where the M_i , $i = 1, 2, \dots, m$ are the class mean vectors and S is the common covariance matrix. For simplicity at this stage, a zero - one loss function and equal prior probabilities of the classes are assumed.

Under these assumptions, the classifier is linear with discriminant functions of the form

$$D_i(X) = W_i^T X + c_i \quad i = 1, 2, \dots, m$$

where

$$W_i = S^{-1} M_i \text{ and } c_i = -\frac{1}{2} M_i^T S^{-1} M_i$$

The covariance matrix is assumed constant, not subject to adaptation. For each observation X to be used for adaptation, the components of the mean vector of class ω_i into which X has been classified is updated according to

$$m'_{ik} = (1-\alpha_k) m_{ik} + \alpha_k x_k, \quad k = 1, 2, \dots, n$$

where α_k is an adaptation parameter which may have different values for different components of the multidimensional measurement space.

However, to avoid the instability problem noted earlier, not every classified vector is used for adaptation. Instead, only those which are "confidently classified" are used, which is accomplished by "thresholding" the discriminant values: an observation X is used to update the mean vector of the class into which it was classified only if the value of the discriminant lies within a specified range, say $D_i \pm T_i$, where D_i is the mean value of the discriminant for class ω_i and T_i is the threshold for class ω_i .

Thus, to completely define the classifier model it must be possible to specify

- (1) the discriminant threshold values, T_i
- (2) the adaptation parameters, α_k .

In selecting the updating vectors, we want to choose vectors that are most likely to have been correctly classified. The basic approach is to select a percentage of the vectors closest to the class mean vectors. Since at any particular time during classification we do not know whether a vector is in the acceptable percentage, we decide this question by considering the distribution of discriminant function values.

We select the $100P_T\%$ "best" classified vectors by finding thresholds T_{1i} and T_{2i} such that

$$\text{Prob } (T_{1i} \leq D_i(X) = W_i^T X + c_i \leq T_{2i}) = P_T$$

This selection procedure is illustrated in Figure 7.1

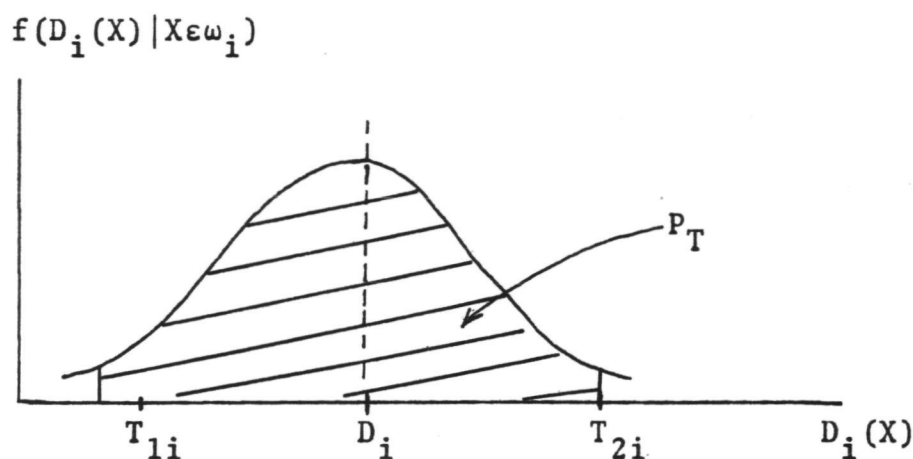


Figure 7.1 Selection of Updating Vectors

Since the class discriminant functions are linear combinations of the mean vector components, assumed to be jointly Gaussian distributed, the distribution of the discriminant function $D_i(X)$ is univariate Gaussian with mean μ_{D_i} and variance $\sigma_{D_i}^2$. These parameters are given by

$$\mu_{D_i} = E[D_i(X) | X \in \omega_i] = W_i^T M_i + c_i$$

$$\sigma_{D_i}^2 = E[(D_i(X) - \mu_{D_i})^2] = M_i^T S^{-1} M_i = -2c_i$$

Therefore the P_T of Eqn. 7.1 is given by

$$P_T = 1 - 2Q \left[\frac{T_i}{\sigma_{D_i}} \right]$$

where

$$Q(Z) = \frac{1}{\sqrt{2\pi}} \int_Z^{\infty} \exp\left(-\frac{1}{2}t^2\right) dt$$

$$T_i = \mu_{D_i} - T_{1_i} = T_{2_i} - \mu_{D_i}$$

If we define $Q^{-1}(\cdot)$ by $Q^{-1}[Q(Z)] = Z$, then

$$T_i = \sigma_{D_i} Q^{-1} \left[\frac{1-P_T}{2} \right]$$

The $Q^{-1}(\cdot)$ function is well known and is tabulated in statistics tables and can also be approximated numerically [3].

If we assume that spatial variation occurs as a function of scan line number, there is no need to update the classifier after every updating sample is selected. Note that four quantities must be recomputed to update the classifier:

- 1) M_i
- 2) $W_i = S^{-1}M_i$
- 3) $c_i = -1/2 M_i^T S^{-1}M_i$
- 4) $T_i = \sigma_{D_i} Q^{-1} \left[\frac{1-P_T}{2} \right]$

One strategy is to update the M_i 's after every updating vector but only update the other 3 quantities after every line.

In choosing α_k we consider two types of errors in estimating the mean vectors:

- 1) Errors due to finite sample size.
- 2) Errors due to spatial variation.

If α_k is too large, the updated mean estimate can be made very poor by relatively few misclassified updating vectors; if α_k is too small, the updating process will not keep up with the spatial variation. It is important in this algorithm to have relatively good mean vector estimates at every decision, because poor estimates will lead to a high misclassification rate, which will in turn increase the number of erroneous updating vectors. Once this sequence is started, the effect may be cumulative, resulting in classifier performance which is severely degraded.

By constraining the error due to finite sample size to be equal to the error due to spatial variation and minimizing the total error with respect to α , it is possible to derive an expression for α as a function of the data characteristics, the number of classes, and the update selector threshold. The details are complicated and may be found in [2].

Experiments with aircraft scanner data have demonstrated that the adaptive classifier can effectively track the data variation and provide performance which is better than that achievable without adaptation [2]. However, in attempting to test the model on satellite data, it was found impossible to locate data (having associated ground observations) with sufficient variability to provide a convincing test. In fact, efforts to locate such data have led to the need to be able to quantify the data variability, and it is felt that significant progress in evaluation of adaptive techniques for remote sensing data analysis may have to await further results in this quantification effort.

[3] C. Hastings, Jr., Approximations for Digital Computers, Princeton, New Jersey, 1955.

7.4 Use of Context

Contextual information in image data can be utilized in a number of ways. Two that have been studied in the present investigation are (1) improved analysis results through sample classification (rather than classification of individual data points), and (2) compression of results storage through object isolation, classification, and coding. The approach is outlined below; theoretical and procedural details appear in [4].

The classification of a multispectral image involves labeling areas of interest in the image. These areas of interest are groups of image points that have been produced by the sensing of objects such as agricultural fields, bodies of water, and cities. One approach to machine classification of images has been to classify each image point separately. Classification algorithms using point-by-point classification methods have been successful in many applications, but in some cases classification accuracy has been undesirably low.

Human photo-interpreters use spatial properties such as texture, size, and shape in image interpretation. The presence of this spatial information in multispectral images suggests that machine classification of multispectral images may be improved if spatial as well as spectral information is used in the classification algorithm.

The classification method presented here is a two step procedure. First, an image is partitioned into blocks or sets of image points. The image partitioning algorithm is designed so that it is likely that each block contains image points from a single

[4] T. V. Robertson, P. H. Swain, and K. S. Fu, "Multispectral Image Partitioning", Information Note 071373, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, 47906.

object of interest. In the second step of the procedure, the blocks are classified. Classifying blocks instead of individual image points allows the measurement and use of texture and other spatial characteristics of objects that are not apparent when single points are classified separately.

The partitioning algorithm divides an image into disjoint rectangles (blocks) such that each area of interest (object) is approximated by a union of blocks. The basic characteristics of the algorithm are described here.

An image I is a set of points in a plane that is surrounded by a closed curve C of finite length. In our discussion we will assume that the image points of I are defined by all the intersection surrounded by C of a set of equally-spaced horizontal and vertical lines in the plane. A subimage of I is an image J such that $J \subseteq I$.

A partition P of an image I is a finite set of images (I_1, I_2, \dots, I_L) such that

$$I = \sum_{i=1}^L I_i$$

and for $j \neq i$,

$$I_j \cap I_i = \emptyset$$

where \emptyset is the empty set. Each $I_j \in P$ will be called a block of P .

The area of an image J will be denoted $|J|$. The size of J is the minimum of the horizontal and vertical extent of J .

A gray-level function $g(\cdot)$ is a function whose domain is an image and whose range is a bounded interval on the real line. We use $g(X)$ to stand for the gray level at a point $X \in I$. For a given X , $g(X)$ will be considered a random variable whose distribution

depends on X . A gray-level vector $G(\cdot)$ is a vector of gray-level functions:

$$G(X) = (g_1(X), g_2(X), \dots, g_N(X))^T,$$

where each $g_i(\cdot)$ is a gray-level function.

Consider an image J . Let $E(\cdot)$ be expected value. We will use the following notation:

$$M_{g_i}(J) = E(g_i(X) | X \in J)$$

$$M_G(J) = \begin{bmatrix} M_{g_1}(J) \\ M_{g_2}(J) \\ \vdots \\ M_{g_N}(J) \end{bmatrix}$$

We call $M_G(J)$ the mean vector of J . Also let

$$S_{g_i}^2(J) = E((g_i(X) - M_{g_i}(X))^2 | X \in J)$$

$$Z_{g_i}^2(J) = E(g_i(X)^2 | X \in J).$$

An image J is G-regular if for any subimage $K \subseteq J$, $M_G(K) = M_G(J)$. A G-regular image is "homogeneous" with respect to G in the sense that the mean values of the gray-level functions $(g_i(\cdot), i=1, 2, \dots, N)$ are constant throughout the image.

A subimage J of I is G -distinct if J is G -regular, and if for any subimage $K \subseteq I$ that is adjacent to J , $K \cup J$ is not G -regular. In other words, a G -distinct subimage is surrounded by subimages with different mean values of the N gray-level functions of G .

A partition P is G -regular if every block of P is G -regular; P is called G -optimal if every block in P is also G -distinct. Note that a G -optimal partition is necessarily G -regular, but a G -regular partition is not G -optimal if some pair of adjacent blocks have the same mean vectors.

The mean test to determine the G -regularity of an image J is carried out as follows: First J is partitioned into two subimages J_1 and J_2 . J is determined to be G -regular if and only if $M_G(J_1) = M_G(J_2)$. In [4] we show that this test makes no errors if the number of image points per unit area is infinite. We also show in [4] that the G -optimal partition P^* is unique.

We assume that the blocks in the G -optimal partition P^* of I , $P^* = (O_1, O_2, \dots, O_M)$, correspond to the objects in I . Therefore a good partition of I is one that closely approximates P^* . We now present a criterion function that is minimized by good partitions.

Consider an arbitrary partition of I , $P = (I_1, I_2, \dots, I_L)$, and a gray-level function $g(\cdot)$. We first define a criterion $V_g(P)$ for the single gray-level function $g(\cdot)$:

$$V_g(P) = \sum_{i=1}^L \frac{|I_i|}{|I|} S_g^2(I_i)$$

-
- [4] T. V. Robertson, P. H. Swain, and K. S. Fu, "Multispectral Image Partitioning", Information Note 071373, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, 47906.

Recall that the $S_g^2(I_i)$'s are the variances of the blocks in the partition P . A block variance tends to be small if the block contains a single object; but a block that overlaps an object boundary which contains several objects will have relatively high variance. Since in the criterion function block variances are weighted by the block areas, $V_g(P)$ will tend to be small when most of the largest blocks contain only a single object; in other words, when P is approximately g -regular. For a gray-level vector $G(\cdot)$ we define

$$V_g(P) = \sum_{j=1}^N V_{g_j}(P) .$$

We also define a partition error

$$\Delta V_g(P) = V_g(P) - V_g(P^*)$$

and

$$\Delta V_G(P) = V_G(P) - V_G(P^*)$$

$$= \sum_{j=1}^N \Delta V_{g_j}(P) .$$

In [4] we show that $V_G(P)$ is a minimum if and only if P is a G -regular partition.

-
- [4] T. V. Robinson, P. H. Swain, and K. S. Fu, "Multispectral Image Partitioning", Information Note 071373, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, 47906.

Figure 7.2 shows a flow chart of the Recursive Image PARTitioning algorithm, which we call RIMPAR continues to subdivide blocks until the block under consideration is either too small or G-regular. The question of G-regularity is decided by the mean test discussed earlier. The specification of which block sizes are too small is handled by a parameter MINSIZE. In [4] we prove the following result: Assuming no errors are made in determining G-regularity, for any $\epsilon > 0$, there are MINSIZE values for which $\Delta V_g(P_f) < \epsilon$, where P_f is a partition of I produced by RIMPAR in a finite number of steps, and I is assumed to have an infinite number of points per unit area.

In practice MINSIZE is useful in resolving ambiguities in object definition: The user of RIMPAR can use MINSIZE to specify whether he wants certain target areas to be considered large textured objects or sets of small, relatively homogeneous objects.

To implement the mean test, several partitions of J are tried. These trial partitions are generated by $(K_D - 1)$ horizontal and $(K_D - 1)$ vertical, equally spaced lines. Here K_D is an integer greater than 1. The trial partition $P_t = (J_1, J_2)$ that yeilds the most improvement in an estimate of the partition criterion function $V_G(\cdot)$ is used to carry out an approximate version of the mean test. In this approximate mean test we use the multivariate T^2 statistical hypothesis test [5] that assumes the gray-levels in J_1 and J_2 are normally distributed, and tests the hypothesis that $M_G(J_1) = M_G(J_2)$.

[5] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Inc., New York, 1958, pp. 108-109.

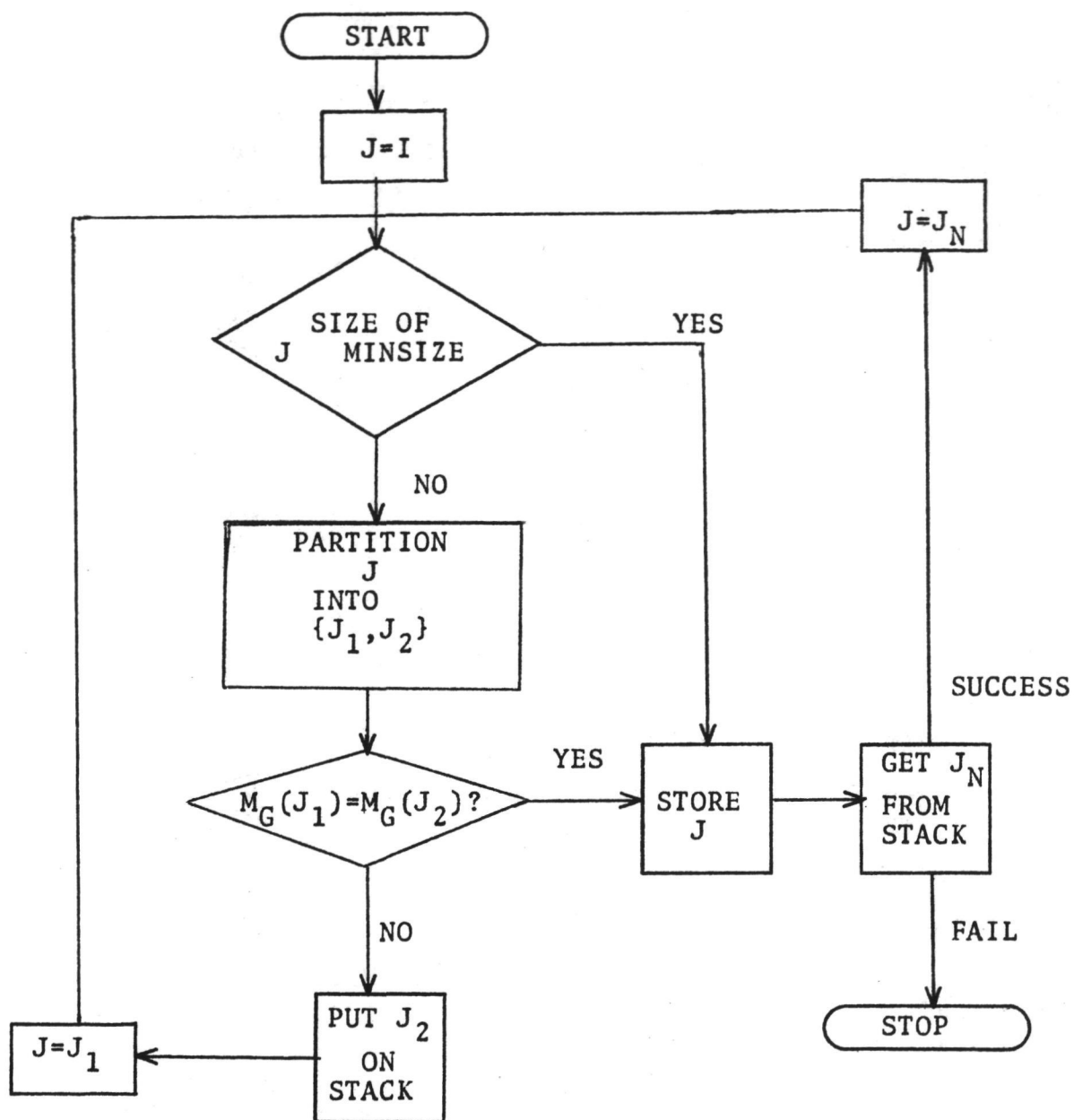


Figure 7.2 Basic RIMPAR Flow Chart

In a sequence of experiments we investigated classifying partitioned images and compared this method to classifying the individual points of images. The classification algorithms used were all based on the assumption that the data are characterized by multivariate normal distributions.

In the supervised classification of partitioned images, a statistical distance measure (the Bhattacharyya distance) was used to determine the distances between the estimated distributions of the gray-levels of subimages of known classification. This technique was compared to supervised per-point classification in which a Bayesian maximum likelihood classifier was used to classify individual image points by comparing point gray-levels to the estimated distributions of the gray-levels of subimages of known classification.

Unsupervised classification was carried out using a standard clustering algorithm, which can be thought of as following these steps:

1. An initial number M of classes is specified, and the initial distributions of these classes are estimated using an arbitrary subset of the data to be clustered.
2. The partition blocks or image points are then classified using supervised classification techniques and the current estimates of the M class distributions.
3. If the class membership of the partition blocks or image points is unchanged from the previous iteration, the algorithm stops.
4. If there is a change in class membership, calculate a new estimate of the M class distributions based on the new members of each class, then return to step 2.

The details of the classification algorithms are discussed in [4].

[4] T. V. Robinson, P. H. Swain, and K. S. Fu, "Multispectral Image Partitioning", Information Note 071373, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, 47906.

In the first set of experiments supervised classification was used to identify crop types in 5 images. The distributions of the classes of interest were estimated before classification using training fields. The characteristics of these 5 images are summarized in Table 7.1. In Table 7.2 we compare RIMPAR classification (classifying an image partitioned by RIMPAR) with per-point classification (classifying individual image points). Classification accuracy is calculated by comparing the classification results with test fields that contain points of known classification. These test fields are distinct from the fields used to estimate distributions used by the classifiers. The processing time reported is in seconds of virtual CPU time on an IBM 360/67 time shared computer. Results storage is in bytes, and is calculated assuming one byte for each class label and 4 bytes to specify a partition block location. The channels used for partitioning and classification are, in general, different for each image. For the aircraft images, wavelengths from 0.40 to 11.7 microns were used, and for the satellite images, wavelengths from 0.6 to 0.8 microns were used.

From the results shown in Table 7.2 we conclude that in comparing per-point and RIMPAR classification, the latter technique gives comparable accuracy (an average of 1% improvement in these experiments), less results storage (24% - 42% in these experiments), and larger processing time (900% - 1250%) compared to the former technique.

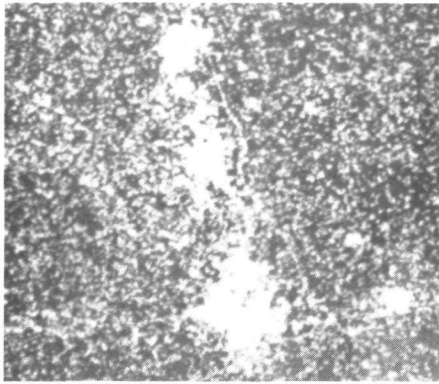
In the next set of experiments, a 93,000 point image from the ERTS-1 satellite was used to investigate the classification of urban areas. This image contains 5 relatively large cities. From top to bottom, the three largest cities are (see Figure 7.3) Jamesville, Wisconsin; Beloit, Wisconsin; and Rockford, Illinois. A smaller city, Belvedere, Illinois appears to the right of Rockford, and above Belvedere is Poplar Grove, Illinois. The goal of these

experiments was to isolate these cities from the rest of the image. This isolation was accomplished by performing unsupervised classification (clustering) of the image and displaying the cluster classes as different gray-levels. The cities were considered to be effectively isolated if they were represented exclusively by a single cluster class. Two methods using clustering were compared: Clustering the individual image points and clustering the partition blocks produced by RIMPAR.

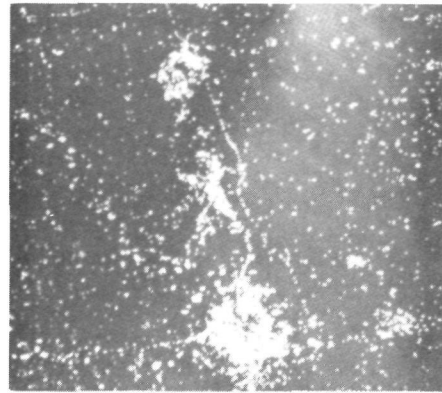
In Figure 7.3 we show the results of clustering the individual points of the image into 5 classes using Bands 5 (0.6 - 0.7 micrometers) and 7 (0.8 - 1.1 micrometers). Visually this clustered image seems to be a good representation of the cities in the image. However, the human visual system does a lot of spatial integration in viewing such a picture. As shown in the right side of Figure 7.3, the cluster class most nearly representing the cities consists of (1) separated points within the cities, and (2) many superfluous points outside the cities. Thus the image description stored in the computer, represented by Figure 7.3 does not specify 5 major objects that represent cities. The cities are not found as distinct objects when individual points are clustered because cities are characterized by texture as well as the reflectance of individual image points.

In Figure 7.4 we show the results of clustering the image using Bands 5 and 7 after the image was first partitioned by RIMPAR. From the figure it is clear that the cities have been approximately isolated. Although the boundaries of the cities are not precise, the image of Figure 7.4 is a useful input to more detailed processing.

In summary, an image partitioning algorithm has been developed and applied to the classification of agricultural and urban areas. This method of classification has been shown to require small classification results storage at the expense of large computation time. The technique has also been shown to be superior to a per-point method in isolating cities in an ERTS-1 image.

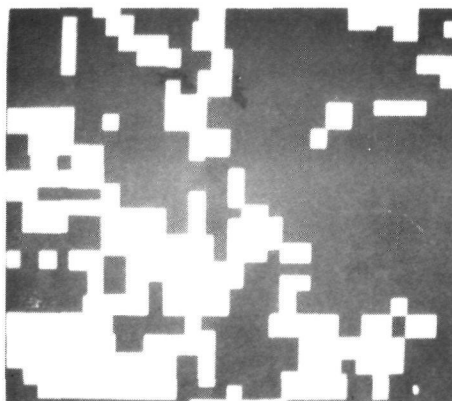


5 Cluster Classes

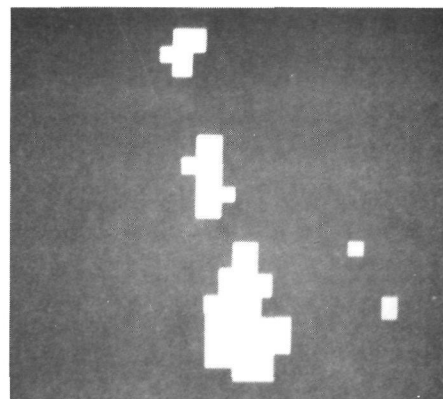


Class 5 Shown as White

Figure 7.3 Per-Point Clustered Satellite Image



5 Cluster Classes



Class 4 Shown as White

Figure 7.4 Clustered Partitioned Image

Table 7.1 Image Characteristics

| Image | Source | Classes of Interest | Average Field Size (Points) | No. Training Field Points | No. Test Field Points |
|----------|------------------------|--|-----------------------------|---------------------------|-----------------------|
| 69002901 | Aircraft 2400 feet | Corn, Soybeans, Wheat Forage, Forest, Water | 221 | 2727 | 5237 |
| 66000600 | Aircraft 2600 feet | Corn, Soybeans, Wheat, Oats, Clover, Alfalfa, Bare Soil | 410 | 4459 | 13562 |
| 71053900 | Aircraft 5000 feet | Corn, Soybeans, Forage, Forest, Water | 64 | 1387 | 6410 |
| 7203280A | Satellite 580 miles | Corn, Soybeans, Other (Other Vegetation) | 18 | 850 | 4842 |
| 7203280B | Satellite 580 miles | Corn, Soybeans, Other (Other Vegetation) | 18 | 1309 | 1409 |

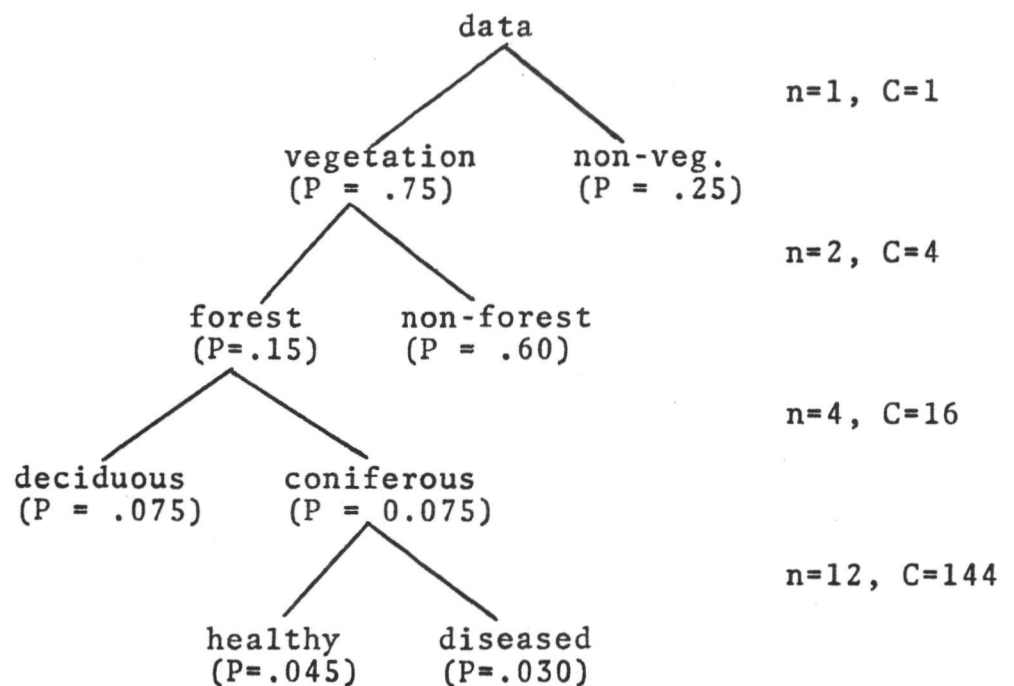
Table 7.2 Comparison of RIMPAC and Per-Point Classifiers

| Image | No. Channels to Partition/Classify | Accuracy* | | Time | | Results Storage | |
|----------|------------------------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | | RIMPAC/Per-Point | RIMPAC/Per-Point | RIMPAC/Per-Point | RIMPAC/Per-Point | RIMPAC/Per-Point | RIMPAC/Per-Point |
| 69002901 | 2/4 | 76.7/78.5 | 76.7/78.5 | 1214/100 | 1214/100 | 15630/44000 | 15630/44000 |
| 66000600 | 2/4 | 79.9/78.5 | 79.9/78.5 | 967/95 | 967/95 | 9535/40280 | 9535/40280 |
| 71053900 | 2/4 | 95.4/93.2 | 95.4/93.2 | 1145/105 | 1145/105 | 13950/46509 | 13950/46509 |
| 7203280A | 2/2 | 82.6/81.3 | 82.6/81.3 | 753/81 | 753/81 | 14125/36000 | 14125/36000 |
| 7203280B | 2/2 | 74.0/71.8 | 74.0/71.8 | 615/67 | 615/67 | 11635/27900 | 11635/27900 |

* Accuracy calculated as 100X (Number of correctly classified points)/(Number of test field points).

7.5 Layered Classifiers

Layered classifier (i.e., multilevel) decision logic provides a capability for making maximal use of available multispectral information at minimal data processing cost. As a simple illustration of the layered classifier concept, consider the following diagram of a hypothetical layered decision structure designed to



classify forested areas and detect diseased coniferous forest. Indicated on the diagram is the *a priori* probability (P) of each class, the number of spectral channels used for each decision layer (n) and the cost (essentially computation time) required for each layer (C). The cost is assumed proportional to the square of the number of channels (which is approximately the case for a Gaussian maximum likelihood procedure).

The average cost of classifying a data point using the layered decision procedure is

$$\bar{C}$$
$$L = (1.00 \times 1) + (.75 \times 4) + (.15 \times 16) + (.075 \times 144) = 17.2$$

If twelve channels were used for classifying every point rather than just for the layer discriminating healthy from diseased conifers, the average cost would be 144, or over eight times as great as for the layered procedure.

Additional motivation for use of layered classifiers is drawn from the following observations:

(1) For problems involving limited training sets for use in classifier design (which is usually the case in practice), inherent dimensionality characteristics may limit the number of features which can be used. That is, the classifier accuracy may actually be better if a subset of the available features is used rather than all of them.

(2) When subsets of the available features are used, the optimal subset for discriminating classes may differ from class to class.

Thus, the advantages of layered classifiers involve both efficiency (cost) and accuracy. The major difficulty in implementing layered classifiers is the difficulty in optimizing the decision tree structure. A very large number of decision trees can be constructed from a given set of classes and features. To seek a decision tree classifier which is general enough to handle classification problems with multiclass and multivariate data sets, two design approaches are here investigated. Experimental verification of these approaches has emphasized problems assuming multivariate normally distributed data sets,

which actually corresponds to a class of multispectral remote sensing classification problems.

To make the discussion clearer, we will first introduce several terms to be used. A "tree" is a graph, each of whose nodes has a unique immediate ascendent node except for a distinguished node, called the "root node", that has no ascendant node. A "terminal node" in a tree does not have descendant nodes; otherwise it is a "non-terminal node". In a "decision tree", a decision is made at a non-terminal node, where the immediate descendant nodes represent the possible decisions. For a "decision tree classifier", an observation is classified by following a path from the root node to a terminal node whose class designation determines to which class the observation belongs.

Ideally, the objective of the decision tree optimization would be to maximize both the classification accuracy and the computation efficiency. However, simultaneous optimization of both accuracy and efficiency is generally impossible, because these two factors are dependent in the sense that one usually has to be sacrificed to some extent to improve the other. Thus for most problems, two types of criterion functions to evaluate the performance of a classifier are considered. One type deals with the total cost, i.e., a combination of accuracy and efficiency; another deals with accuracy only (which can be considered a special form of the first type).

7.5 1 Design for Maximal Accuracy: Binary Decision Trees.

In a binary decision tree, each non-terminal node has exactly two immediate descendant nodes. For our purposes this corresponds to a test of likelihood for a pair of classes, using their optimal feature subset.

An illustration of the binary tree procedure is shown in Figure 7.5 for classifying an unknown into four classes $\{\omega_1, \omega_2, \omega_3, \omega_4\}$. In this figure the class of a terminal node is the final decision. Let $F(i,j)$ denote the optimal feature subset used in the decision function for classifying classes ω_i and ω_j . In this figure the class of a terminal node is the final decision. Let $F(i,j)$ denote the optimal feature subset used in the decision function for classifying classes ω_i and ω_j .

For n -class classification $n-1$ tests are necessary to reach a terminal decision. In an optimal binary tree procedure, to reach a terminal decision for n -class classification, a sequence of $n-1$ tests are performed; in each test a Bayesian decision rule is used to classify a pair of classes (i.e., to discriminate one class from another), and the class rejected in the test is excluded from consideration in further tests.

The mathematical formulation of the binary tree procedure is as follows:

Assuming D is the optimal decision function for classifying class pair ω_i and ω_j , and \mathcal{D} is the decision of D , we have

$$\mathcal{D} = D(\omega_i, \omega_j)$$

with

$$\mathcal{D} = \begin{cases} \omega_i & \text{if } r_{ij} \geq 1 \\ \omega_j & \text{otherwise} \end{cases}$$

where

$$r_{ij} = \frac{p(x|\omega_i)}{p(x|\omega_j)}$$

is the likelihood ratio for two classes ω_i and ω_j . With \mathcal{D} and D

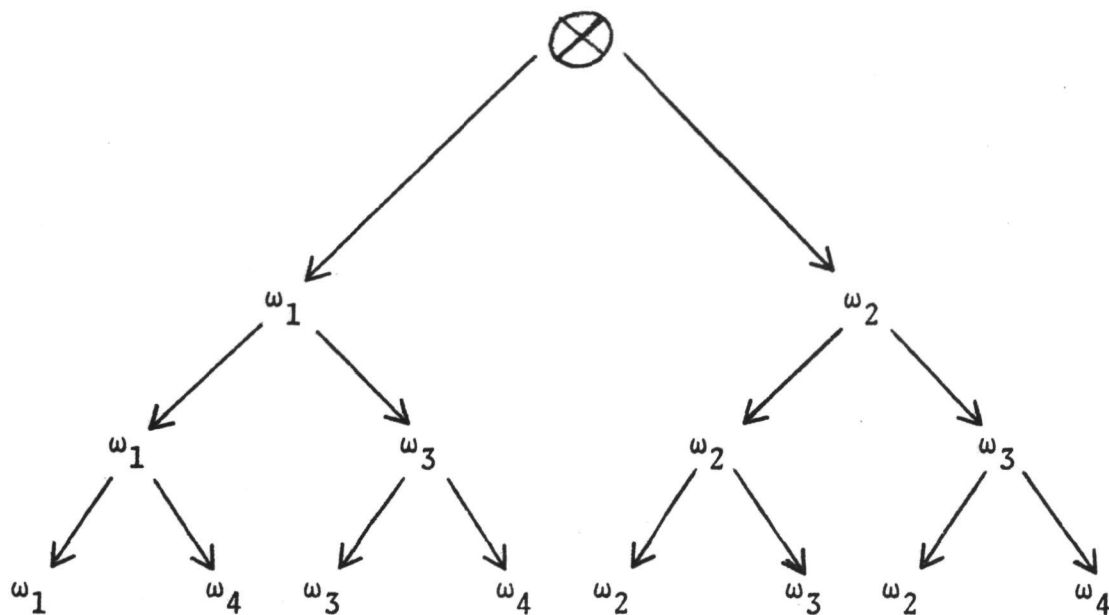


Figure 7.5 A binary decision tree for four class classification.

defined as above, the binary tree procedure can be put in a recursive form:

$$\mathcal{D}_n = D(\omega_n, \mathcal{D}_{n-1})$$

$$\text{with } \mathcal{D}_1 = \omega_1$$

where n is the number of classes, \mathcal{D}_n is the final decision.

Since each unknown is classified into a class through $n-1$ tests, it is not necessary to construct and store the entire tree structure shown in Figure 7.5. If the densities of all classes can be estimated, the necessary information to construct the binary tree decision procedure (as described in Equation (7.4) to Equation (7.7)) would be to use the optimal feature subsets for all class pairs. After the optimal feature subsets for all class pairs are found by feature selection techniques, the remaining decision procedure is shown by the block diagram in Figure 7.6.

Some experimental results for the binary tree procedure will be presented here. Data used for classification were multispectral data gathered by a multispectral scanner with twelve spectral bands. The dimensionality of each data vector (the number of available features) was therefore twelve. In the first experiment, data sets totalling 4,636 samples from five classes were used. Approximately ten percent were used to estimate the probability distributions (approximated by multivariate normal distributions), and all were used to test the classification accuracy. Both the Bhattacharyya Distance and the Divergence were tested as separability criteria for feature selection. Classifiers with dimensionality (for each test) of three, four and five were constructed. The classification results are listed in Table 7.3, together with the

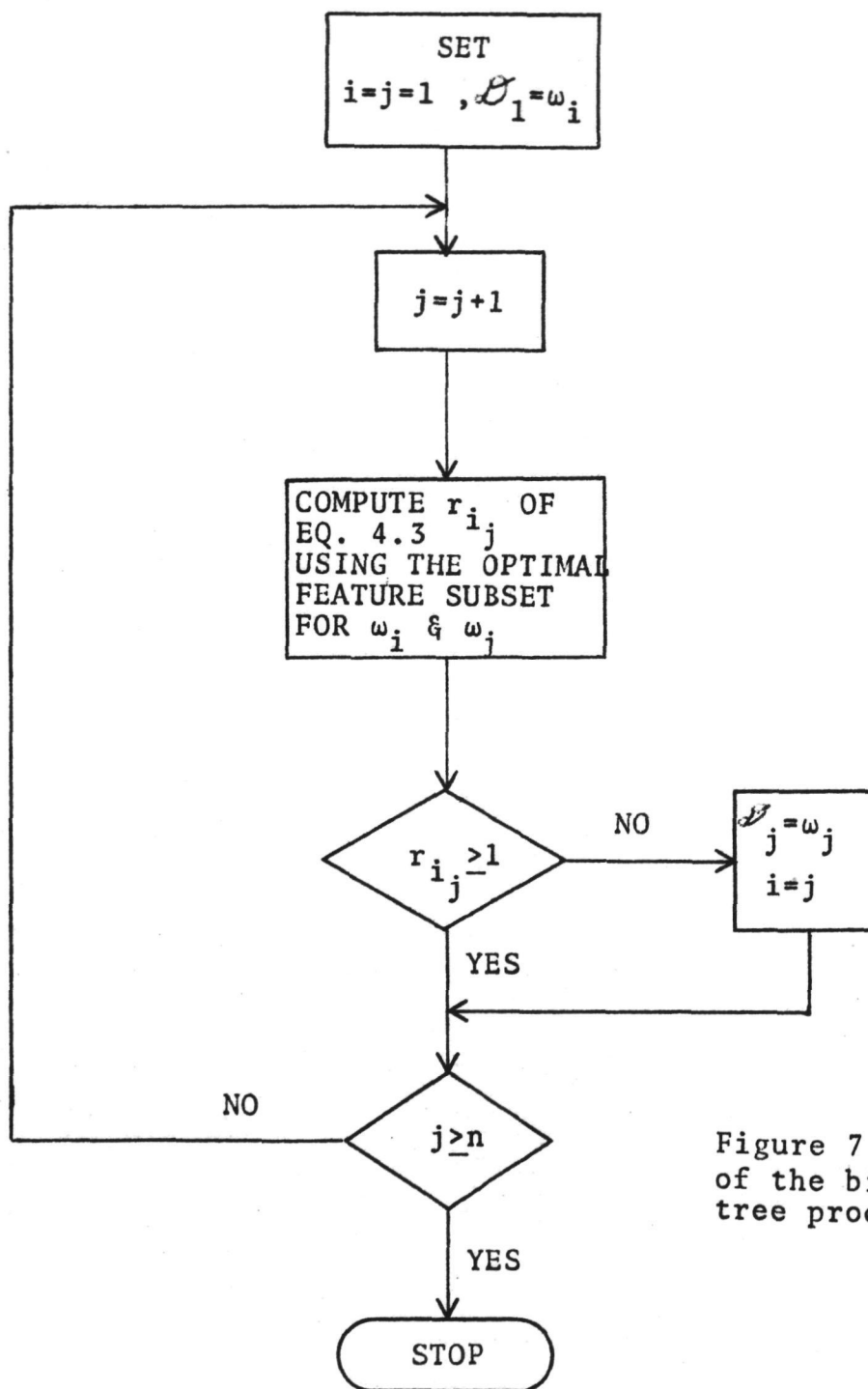


Figure 7.6 Flow chart of the binary decision tree procedure

| DESCRIPTION | MAXIMUM LIKELIHOOD PROCEDURE | | | BINARY DECISION TREE PROCEDURE | | |
|----------------|---------------------------------|----------------|---------|-----------------------------------|----------------|--------|
| | AVERAGE | AVERAGE | BEST | DYNAMIC PROG | | SEARCH |
| DIMENSIONALITY | B _T | D _T | RESULTS | B _T | D _T | |
| 3 | 18.1 | 22.8 | 18.1 | 21.1 | 21.4 | 17.7 |
| 4 | 18.5 | 20.2 | 18.5 | 17.8 | 20.8 | 18.3 |
| 5 | 20.3 | 20.3 | 18.7 | 18.2 | 19.9 | 20.6 |

Table 7.3 Classification Results (% Error) of Conventional and Binary Tree Procedures of Experiment I.

| DESCRIPTION | CONVENTIONAL PROCEDURE | | BINARY TREE PROCEDURE | |
|----------------|---------------------------|----------------|--------------------------|----------------|
| | AVERAGE | AVERAGE | DYNAMIC PROG | DYNAMIC PROG |
| DIMENSIONALITY | B _T | D _T | B _T | D _T |
| 3 | 22.8 | 18.0 | 6.7 | 8.2 |
| 4 | 8.1 | 8.0 | 7.0 | 7.2 |
| 5 | 7.5 | 7.6 | 6.7 | 6.7 |

Table 7.4 Classification Results (% Error) of Experiment II.

results using conventional procedures with maximum likelihood decision rule. In the latter case, feature subsets were selected according to maximum average transformed Bhattacharyya Distance B_T , and maximum average transformed Divergence D_T .

Results of binary decision tree procedures designed by selecting feature subsets from a set of "likely" feature subsets (feature subsets with high average B_T) with separability criterion B_T are also listed in Table 7.3. The best results obtained for the conventional procedure (by testing several highly ranked feature subsets of same dimensionality) and the binary tree procedure are plotted in Figure 7.7. Notice the optimal dimensionality for the conventional procedure is three. A binary tree with this dimensionality does achieve the highest accuracy, which is higher than that achieved by the conventional maximum likelihood procedure.

In the second experiment, the same procedure as in the first experiment was used, except nine classes of fairly separable data sets (totalling 4894 samples, approximately one-fifth used for training) were selected. The classification results are listed in Table 7.4; results of conventional and binary tree procedures with B_T as feature selection criterion are plotted in Figure 7.8.

From the results of the above two experiments, it is observed that binary tree procedures can provide better performance than the conventional procedures. In both experiments, maximum accuracies have been achieved by the binary tree procedures with feature dimensionality in each test being less than that of the complete set. The efficiency of the binary tree procedure is generally lower than that of a conventional procedure using the same feature dimensionality, because more conditional

- : Results of Binary Tree Classifiers
- : Results of Conventional Classifiers
- : Best Results of that Dimensionality

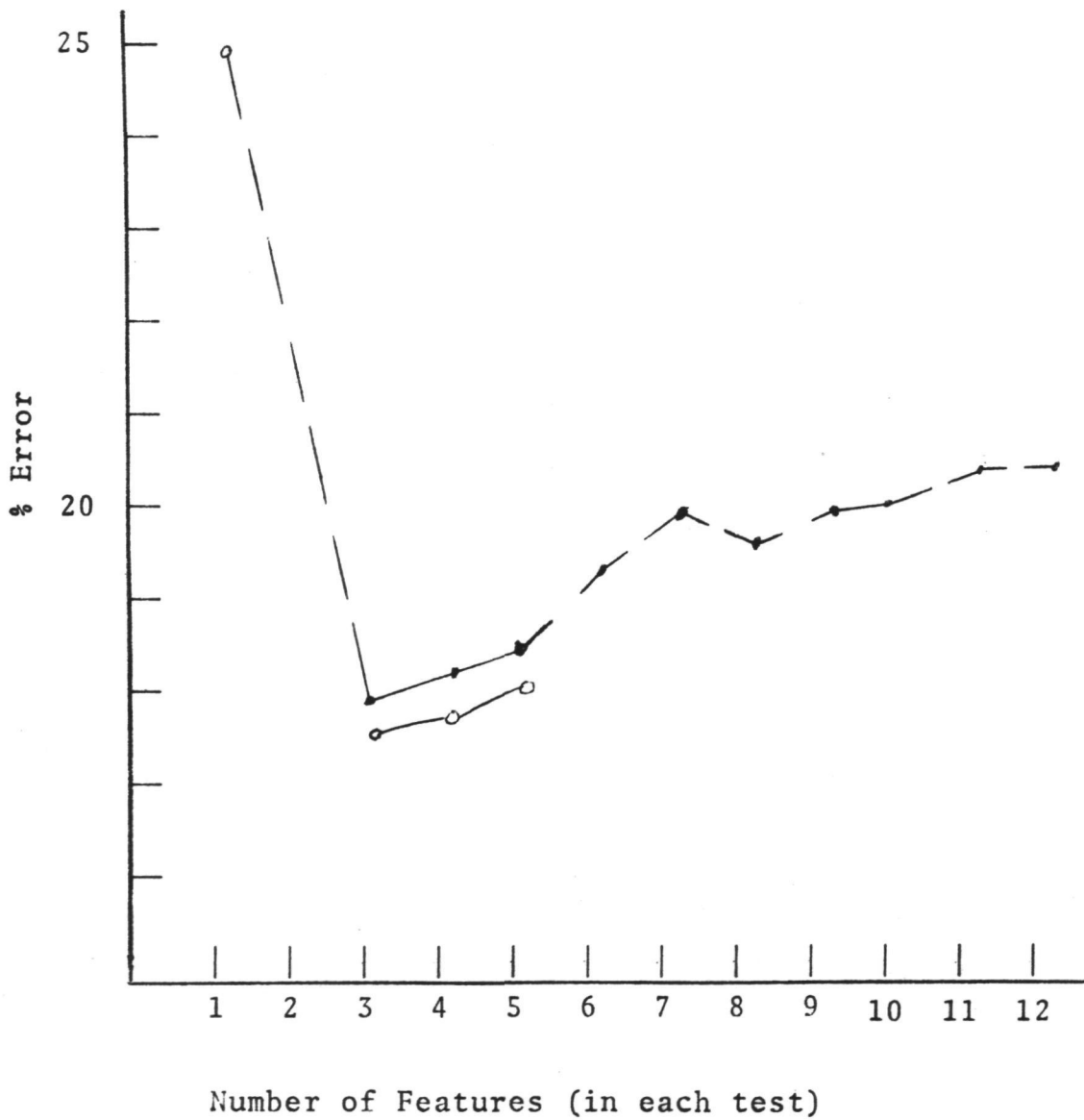


Figure 7.7

—○—: Results of Binary Tree Classifiers
—●—: Results of Conventional Classifiers

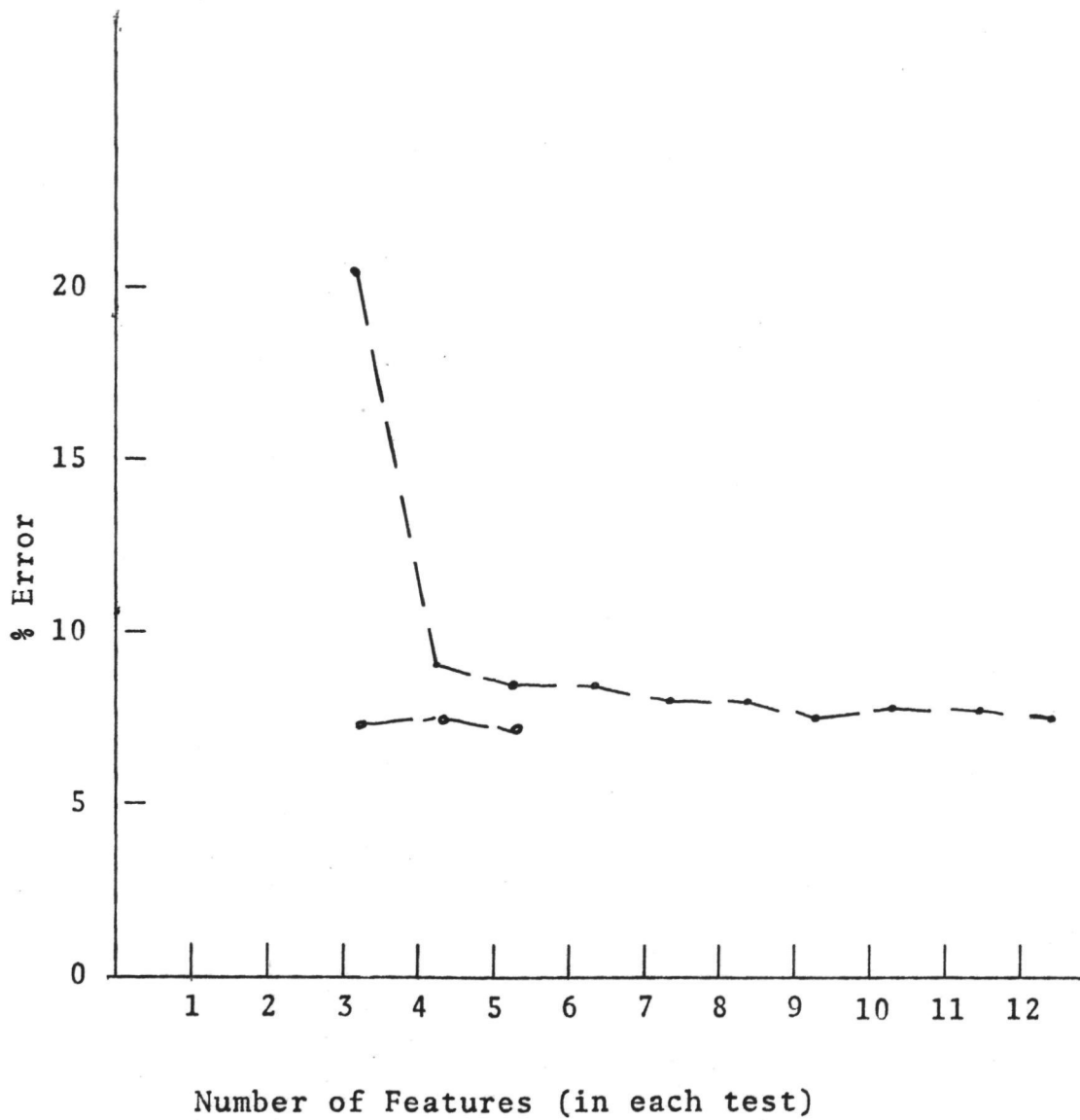


Figure 7.8

probabilities have to be calculated. If the feature subsets used for all class pairs are different, the number of conditional probabilities calculated is twice that normally calculated using a single feature subset.

A drawback of this method is that after a class fails the comparison test against another class, it is immediately rejected instead of being compared to the rest of the classes. This does not create problems when the same feature subset is used for all $n-1$ tests. When different feature subsets are used, conditional probabilities for different classes are not compared on an equivalent basis. Contradiction of results of classification might occur if the sequence of classes used in tests is different from the sequence $\{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}$ used in Equation (7.7). A different class sequence in the tests corresponds to a different tree structure. For example, the sequence $\{\omega_4, \omega_3, \omega_2, \omega_1\}$ will lead to the structure shown in Figure 7.9, which is an alternative to the structure shown in Figure 7.7. The different results for alternative structures is illustrated by a simple example in Figure 7.10, where the region $(x < 0, y < 0, z > 0)$ in feature space will be assigned to two different classes due to two different arrangements as shown. From this standpoint, it is clear that the binary tree procedure may be sub-optimal with respect to maximizing the accuracy for multiclass classification. But if the probabilities are fairly well represented by the training samples, the sample population in the regions of ambiguity in feature space is very small; therefore the difference in classification results due to different arrangements is negligible.

7.52 The Search Approach to Decision Tree Optimization. For the purpose of "overall" optimization, a tree structure is designed as generally as possible. In particular:

- 1) Any feature subset can be used in the decision function

of a non-terminal node.

- 2) The number of immediate descendant nodes of a non-terminal node varies from two to the number of classes in that node.
- 3) The number of classes in a node is always greater than the number of classes in each of its immediate descendant nodes.
- 4) No two immediate descendant nodes of a non-terminal node contain the same set of classes.

With such generality, the tree classifier designed can no longer be expressed in simple mathematical form as given for the binary tree procedure. Essentially two kinds of information are involved in specifying the decision tree structure. One is the node information which tells how the terminal and non-terminal nodes are linked. The other is the decision function information (the feature subset to be used). The node information can be specified by a string which is a breadth first coding of all the nodes of a tree. Each node is represented by a symbol: the terminal nodes by a unique symbol and the non-terminal nodes by a number equal to the number of its immediate descendent nodes. An example is shown in Figure 7.11. A one-one correspondence exists between a tree structure and its coding string "S".

7.53 The Search Procedure. There are basically two problems in determining the optimal decision tree structure. One is the potential complexity of a tree structure. It is by no means easy to describe the tree structure in terms of a set of variables corresponding to a space in which each point stands for a unique tree structure. The second problem is that the overall performance of a candidate classifier structure cannot be predicted exactly. Because of the first problem, most of the existing mathematical programming procedures cannot be applied effectively. Hence, a heuristic search procedure has been developed in which

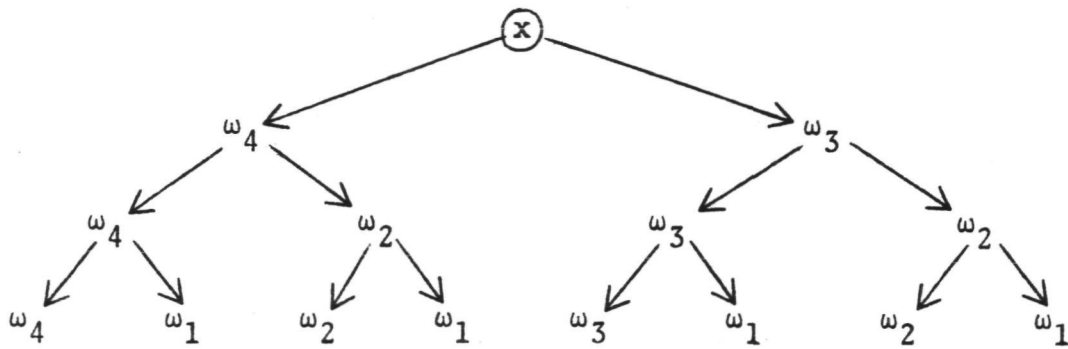


Figure 7.9 Another Binary Decision Tree for Four Class Classification.

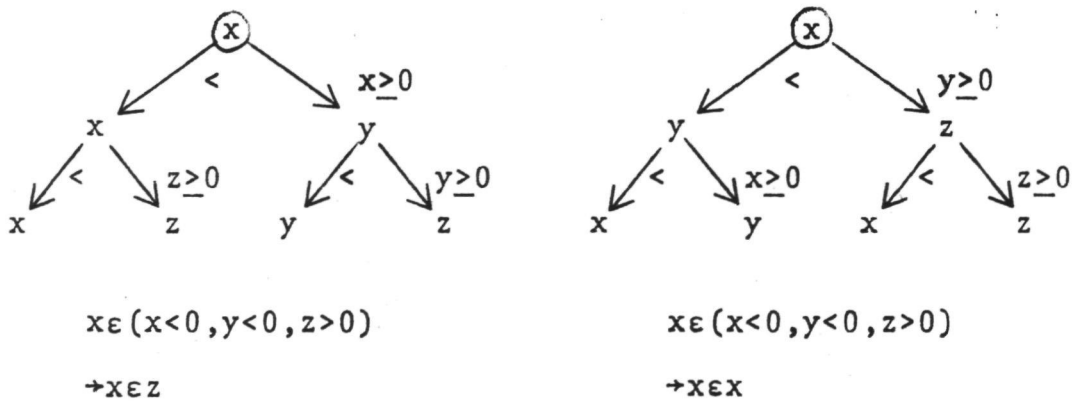
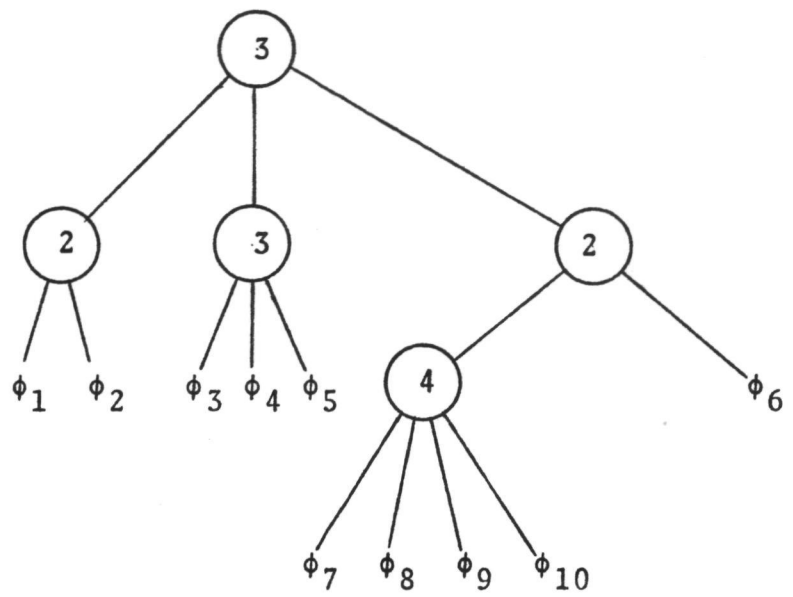


Figure 7.10 Different Results Due to Different Tree Structures.



$S = 3 \ 2 \ 3 \ 2 \ \phi_1 \ \phi_2 \ \phi_3 \ \phi_4 \ \phi_5 \ 4 \ \phi_6 \ \phi_7 \ \phi_8 \ \phi_9 \ \phi_{10}$

Figure 7.11 A Tree Structure with its String

the decision tree is constructed stage by stage. For the second problem, there is no direct solution, but means are available for estimating the performance with reasonable accuracy.

This search procedure first selects a set of feature subsets to be considered. If m , the total number of features is small, all $2^m - 1$ feature combinations can be used. If m is large, feature selection methods can be used to select a set of "likely" feature subsets out of the $2^m - 1$ possibilities. The reduction in feature subsets increases the search efficiency.

The selected feature subsets are then searched in order to construct a stage of the decision tree structure. For each feature subset and the classes under consideration, a nonsupervised clustering is performed based on the class separability for that feature subset and a candidate sub-structure (a stage of the tree) is constructed. All candidate sub-structures are then evaluated using a function which reflects the cost of classification at that stage, and the best of the candidates is selected. The corresponding feature subset is used for classification; the statistical parameters are the pooled statistics of representative classes in each group.

After a stage of the decision tree is constructed, some newly generated nodes may have more than one class. The same procedure is used in expanding those nodes, i.e., constructing the next stages. The search procedure terminates, the decision design completed, when all new nodes contain only one class.

A flow chart for the search method is shown in Figure 7.12. In the search procedure, once the feature subsets for search are selected, the feature clustering and evaluation are the two most important steps. The non-supervised clustering method is described in detail in reference [6].

[6] C. L. Wu, P. H. Swain, and D. A. Landgrebe, "The Decision Tree Approach to Classification", Information Note 090174, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, September 1974.

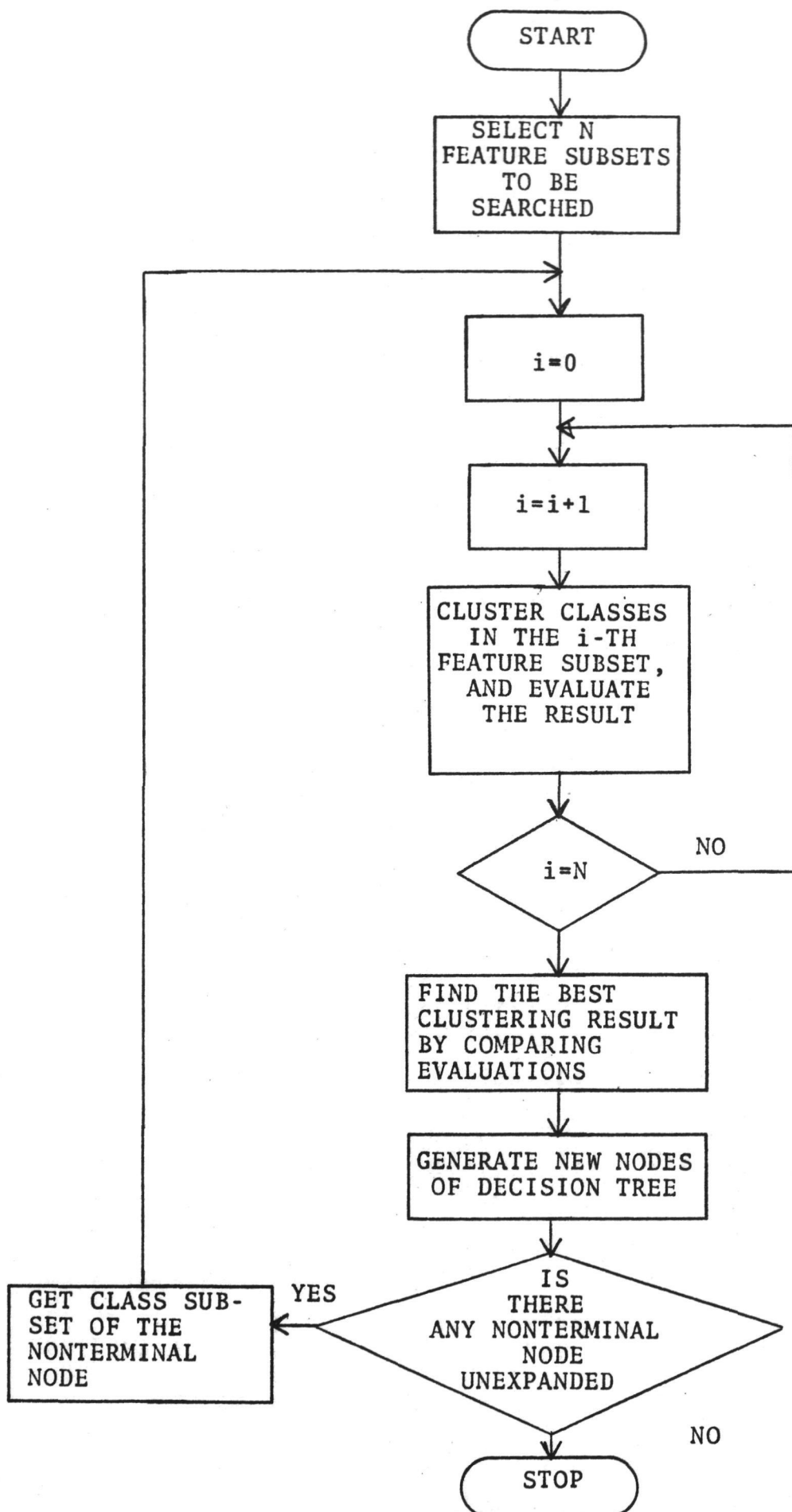


Figure 7.12
Flow chart of
search
procedure

7.54 The Evaluation Function. Searching for the "best" decision tree requires a means for evaluating each candidate. The functional form used also defines what is meant by "overall performance", in terms of accuracy and efficiency. The form which has been selected for this study is a weighted sum. In particular the evaluation of the decision function for each candidate structure *following* node d_i has the form:

$$E(d_i) = -T(d_i) - K \mathcal{E}(d_i) + \sum_{j=1}^C E(d_{i+j})$$

The first two terms give the contribution due to node d_i ; the summation is the estimated evaluation of succeeding stages. $T(d_i)$ is computation time, and $\mathcal{E}(d_i)$ is classification error. K is the weighting constant, specified by the designer, which determines the relative importance of efficiency and accuracy.

The evaluation function is simple but its application in practice is quite a different matter, complicated by such factors as the unavailability of a direct predictor of classification accuracy (needed to compute the $\mathcal{E}(d_i)$ term. The details are too complicated to present in this report, but may be found in reference [6].

The search heuristic produces decision tree structures which are likely to be suboptimal for a number of reasons, including

- 1) Only a subset of all possible trees are actually considered.
- 2) A number of approximations are made in the process of computing the evaluation function.

[6] C. L. Wu, P. H. Swain, and D. A. Landgrebe, "The Decision Tree Approach to Classification", Information Note 090174, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, September 1974.

However, it is generally the case that a large number of "nearly optimal" solutions are possible which can produce results so close to the best obtainable that the additional cost of finding the optimal design is not warranted. The search heuristic is designed to produce either the optimal design or one of the nearly optimal designs with high probability. The experimental results discussed below demonstrate this capability.

In each of the following experiments, different classifier structures were produced by varying the design parameters and options, including:

- 1) The maximum number of features used in each decision stage.
- 2) The distance criterion used in the search procedure.
- 3) The threshold value used to determine class associativity in the non-supervised clustering.
- 4) The constant K which determines the relative importance of accuracy to efficiency.

The first experiment was conducted on data sets used in the second experiment with binary decision trees, where nine classes were to be classified. A number of different decision tree classifiers were designed using the search procedure with various options as described above. The results for the decision tree classifiers designed are plotted in the upper part of Figure 7.13 in terms of "time-ratio" and percent error. ("Time-ratio" is the ratio of total time to the time required for the conventional classifier using four features.) The detailed description of the classifiers and the numerical results are tabulated in reference [6].

It is noticed that for any given computation time, the error of the decision tree classifier can be lower than that of the conventional classifier. Or, for a given level of accuracy the computing time can be reduced by using properly designed decision trees.

The probability P_i that a classification path will pass through node d_i has been approximated during the design process. As a result the expected amount of computation time per sample for a given design can be approximated by properly summing up the products of probabilities and computation times for all stages. This expected value can then be compared to the true computation time. Since the a priori probabilities of classes are usually unknown or roughly estimated, the assumption of equal a priori probabilities has been used in designing the decision tree classifiers. To test the accuracy of the time estimation under the assumption of equal a priori probabilities, and to provide additional results of the search procedure, simulated data sets were classified having the same distributions as the real data sets used in the previous tests but normally distributed with 1,000 samples per class. The results are shown in the lower part of Figure 7.13. And the expected computation times vs. the measured computation times (in relative units) are plotted in Figure 7.14. The accuracies for the simulated data being higher than those for the real data was to be expected because the real data are only approximately normally distributed. The nearly constant accuracy and the closeness of the approximated and measured classification times indicate that the search procedure is performing as desired.

A second experiment was conducted on ERTS multispectral scanner data, with four spectral bands. Twenty-six spectral classes were found from training areas, belonging to five

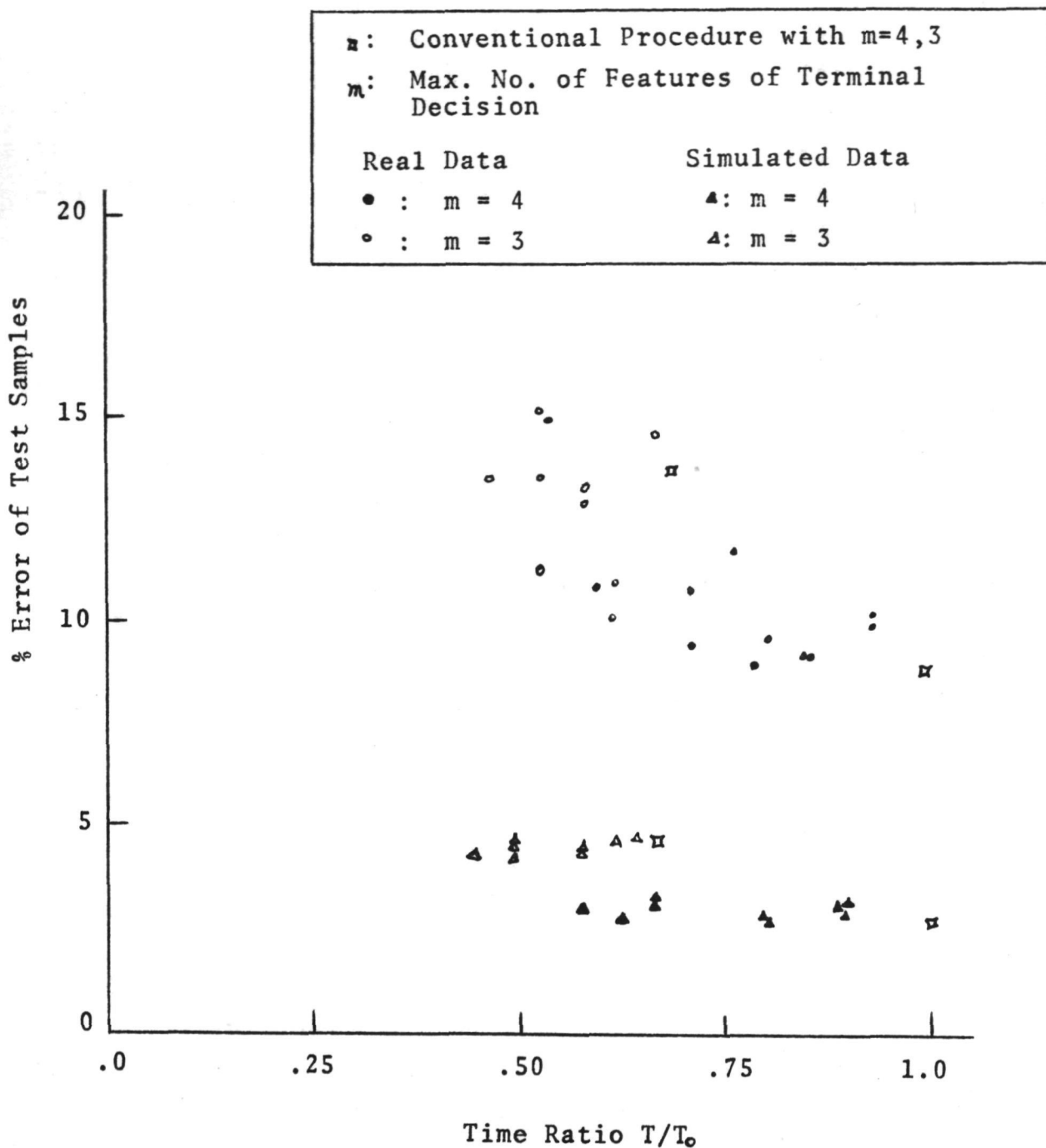


Figure 7.13 Performance of Decision Tree Classifiers on Real and Simulated Data Sets (T = time required by conventional classifier using .4 features)

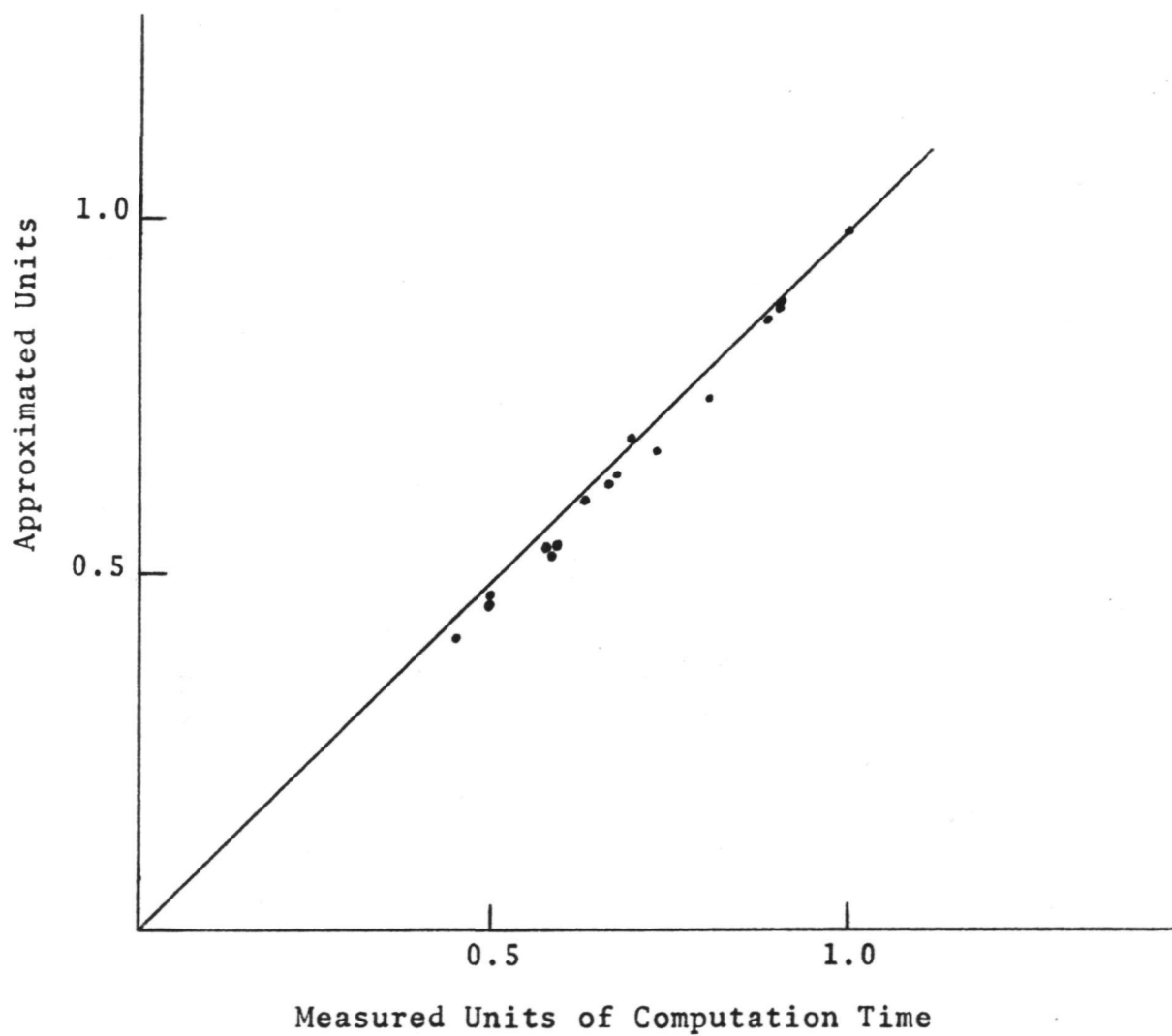


Figure 7.14 Predicted and Measured Computation Time

meaningful groups. The test area to be classified contained 12,467 sample points, 773 samples of known ground cover type. The latter were used to estimate the classification error rate. Classification results for different decision tree classifiers designed are plotted in Figure 7.15a and Figure 7.15b for distance criteria B_T and D_T , respectively. A typical decision tree classifier designed is shown in Figure 7.16, where numbers indicate the spectral classes. The tree structure shown in the lower part of Figure 7.16 is a duplicate of the tree above it except that each symbol indicates the group that class belongs to. Again, it is noticed that the efficiency of classification has increased but in this case with no appreciable increase in error rate. Indeed the most accurate results in this experiment were achieved by decision tree classifiers.

7.55 Summary. The design of optimal decision tree classifiers is not a simple problem. The procedure for selecting an optimal classifier can be very complex, not only because an enormous number of classifiers can be constructed for any given problem, but also because it is difficult to accurately predict the performance of a classifier. The two approaches taken in this investigation represent steps toward solving the decision tree design problem. For the first approach, because of the relatively simple classifier structure the design procedure is not very complicated. The key step is to find the optimal feature subset for each pair of classes. For the second approach, the design procedure is extremely complicated. Due to the lack of methods to exactly predict the classification probability, several empirical methods have been incorporated in the search procedure. The result is a capability to arrive at decision tree classifier designs, but the designs are at best sub-optimal.

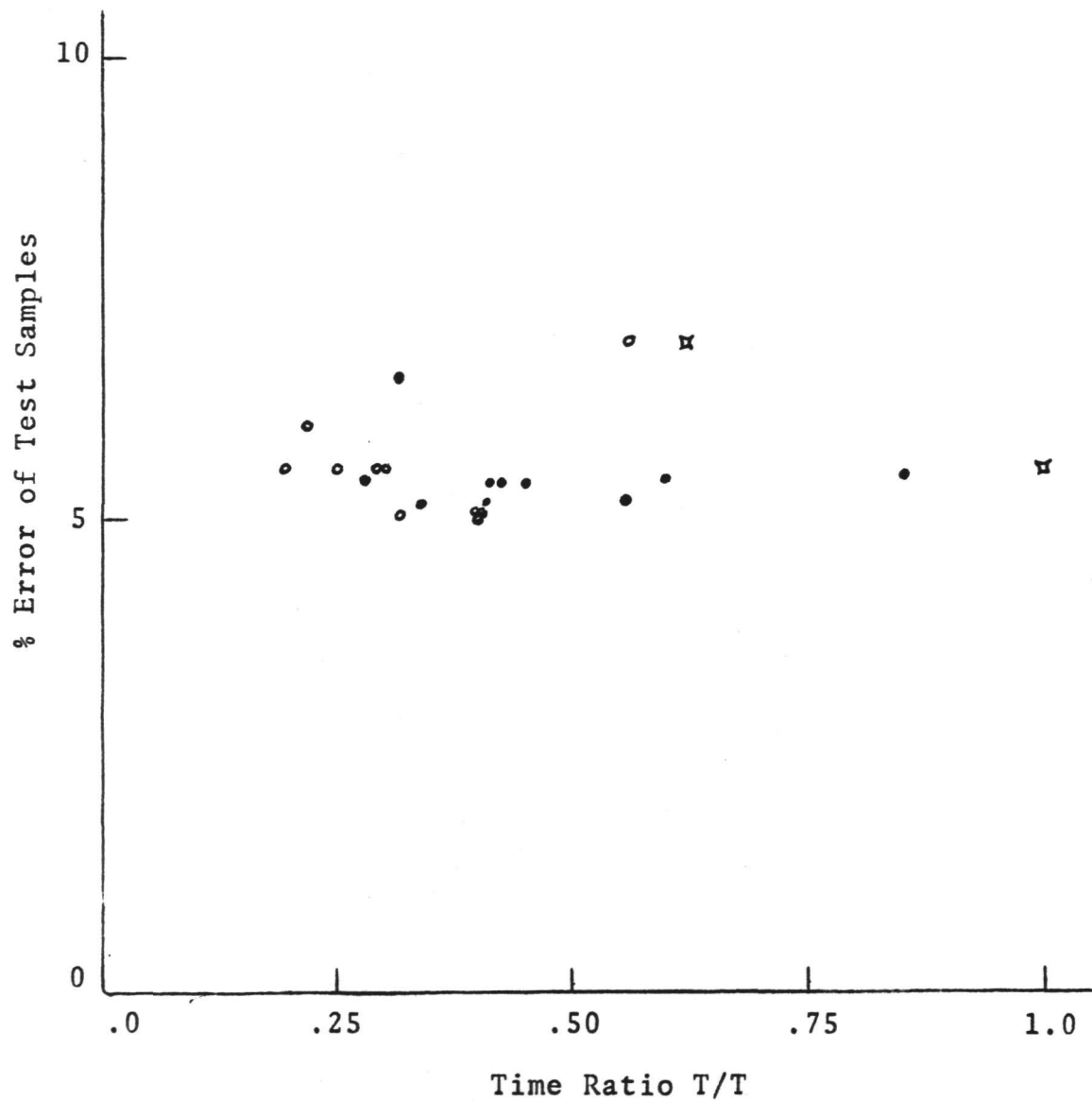


Figure 7.15a Performance of Decision Tree Classifiers Designed with B_T as Distance Criterion

•: $m=4$, ○: $m=3$,

✕: Conventional Procedure with $m=3,4$

Experiments conducted on real data have tested both design approaches. From the results, the original hypothesis that performance of decision tree classifiers can be better than conventional classifiers has been confirmed.

7.6 Analysis Technique Development: Concluding Remarks

Our most conclusive finding in this investigation has been that the multispectral data analysis techniques heretofore developed for aircraft data and, to a much lesser extent, digitized space photography can be effectively applied, with appropriate modifications, to multispectral scanner data from satellites. Most important of the modifications has been upgrading of the cluster analysis capability, which is used in both supervised and nonsupervised analysis modes.

A model has been developed for adaptive classification where large geographical areas are involved. However, the results of attempts to evaluate this model have been inconclusive, largely due to the unavailability of suitable test data.

A method for "recursive image partitioning" has been implemented which utilizes scene context to decompose the scene into self-defined "objects". This technique appears to be of greatest potential use for reducing analysis results storage, although other applications related to unsupervised image analysis can also be envisioned. In its present form, however, the algorithm is computationally very expensive.

The application of layered decision logic has been demonstrated to be a potentially powerful tool for the design of future pattern classifier systems. This approach offers only improved efficiency (speed), but also makes optimal use of available information, within the constraints of inherent data dimensionality, to maximize classification accuracy. Under this contract, promising research results have been obtained and rudimentary software developed. This is a top priority area for further work.

8.0 Data Reformatting and Temporal Overlay

8.1 Introduction

The data reformatting and overlay project was included in the study primarily as a supporting technology task with little emphasis on development of new technology. As the study progressed certain advancements were made as a by-product of application of existing techniques to large volumes of ERTS-1 data. Thus, the task was productive in the technology area as well as enabling access to very large volumes of digital data by the other eight projects. A data handling and reformatting system was developed which included new software and cataloging procedures for the ERTS System Processed CCT data and over 300 frames of CCT data were handled by the system over the course of the study.

In addition to basic data handling functions the project plan included temporal registration of sequential passes over the same area to enable study of time varying effects. Image registration was requested by a large number of investigators over the period of the study and a total of 76 pairs of ERTS-1 subframes were digitally registered during the study.

A data quality evaluation task was included in the reformatting effort to evaluate certain problems in the CCT data. The major problem observed was the so-called "striping effect" which occurred in certain channels of a large number of frames over the 22 months of our utilization of CCT data. The problem was evidently due to calibration problems for the six separate detectors for each band. Other problems such as saturation of data from clouds and snow cover were also investigated. This activity was limited to analysis only since resources were not available to implement corrections in the present contract.

The fourth activity in this project was geometric correction of ERTS-1 data. An existing program was modified to enable an approximate correction of System Corrected CCT data to remove

scale distortion and skew due to earth rotation plus rotate the data to a North-oriented format. This process was not originally planned for; however, many investigators found it extremely difficult to locate areas of interest in computer line printer pictorial reproductions of the ERTS data and a strong need existed for a correction which would make the line printer images match topographic maps of the same area. A theoretical linear transformation was developed which enabled production of North-oriented images at an approximate scale of 1:24000 on the line printer with a scale error of nominally 1% without the use of ground control. By the end of the study, 197 subframes of CCT data were corrected using this capability.

These four activity areas are described in detail in the following sections. Section 8.1 describes the reformatting system, Section 8.2 the data quality analysis that was carried out, Section 8.3 describes the temporal overlay algorithm which was applied to the CCT data and presents some evaluation examples. Section 8.4 describes the geometric correction operations that were developed and made available to investigators.

8.20 Data Reformatting

8.21 Planned Data Flow

The flow of ERTS data from the ERTS space platform to the LARS analyst is shown in Figure 8.1. ERTS project principal investigators at LARS have placed Standing Orders with NASA for their respective test sites. The NASA Data Processing Facility automatically sends LARS black-and-white images of system corrected data as they are collected over the several test sites. These images received via the Standing Order are filed with the LARS Data Coordinator. The researchers responsible for analyzing respective test sites are notified of images received and may check them out for evaluation. If upon examining one or several images of a test site, a researcher decides he requires additional photographic images and/or computer compatible tapes of a particular satellite observation, he makes his requirements known to the respective test site principal investigator, and initiates a NASA Data Request. The Data Request is completed and sent to NASA by the LARS Data Coordinator. When the requested data is received at LARS, it is logged into the LARS data library. Photographic data is filed in the Photographic Data Library and digital tape data is filed in the LARS computer tape library. Again, photographic images may be checked out from the Photographic Data Library. ERTS images received in the form of digital magnetic tapes are not recorded in a format compatible with the LARSYS analysis computer programs and may not be used directly by ERTS data analyst. The digital data becomes available in the form of LARSYS Multispectral Image Storage Tapes upon special request to the Data Reformatting Operations Group.

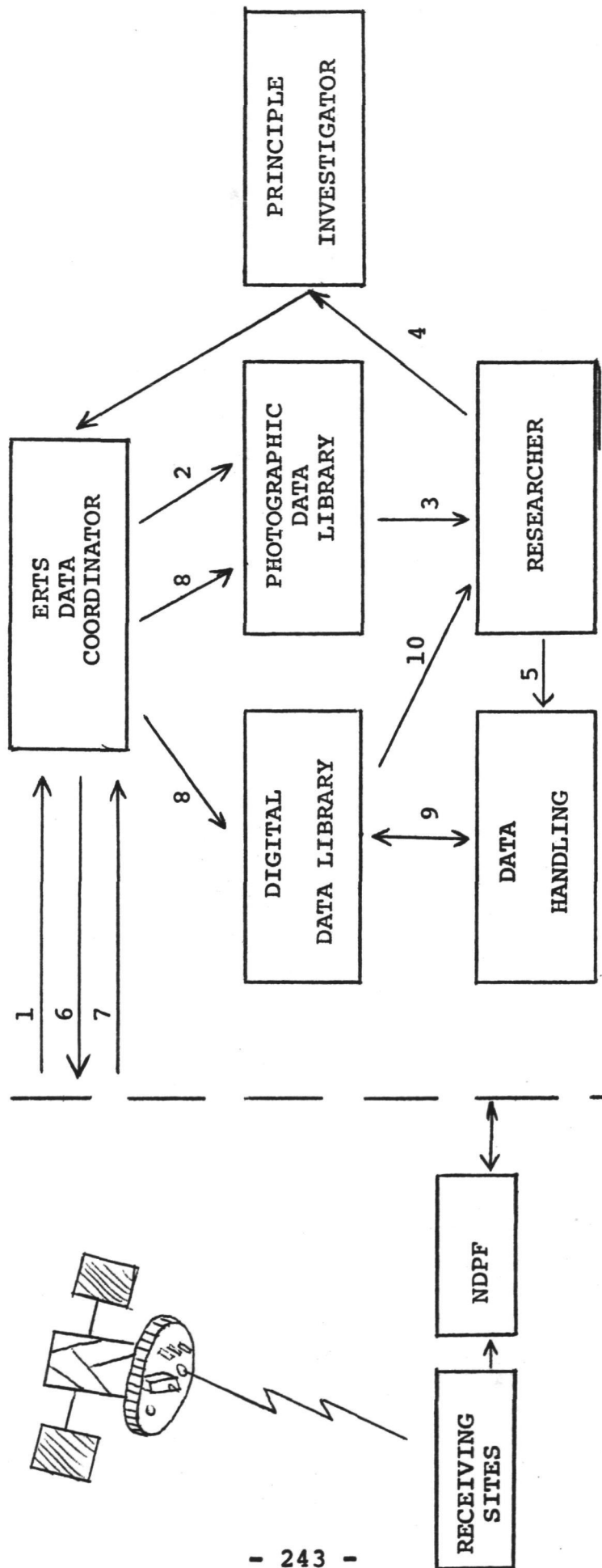
When ERTS computer compatible tapes (CCT's) are checked into the LARS computer tape library, one CCT of each frame is computer verified and the frame annotation and identification records are stored. The stored frame ID data is used to maintain a computer

ERTS

DATA FLOW

NASA

LARS



1. Bulk images via Standing Order
2. Bulk images filled in Photo Library
3. Researcher checks out images
4. Data Request Form initiated
5. Data reformatting requested

6. Data Request submitted to NDPF
7. Specially processed digital and other data received
8. Data is filled in appropriate library
9. Digital data converted to LARSYS format
10. Researcher analyzes LARSYS digital tapes

Figure 8.1

printed catalog containing the frame ID information of all ERTS CCT data checked in. Each page of the catalog contains all ID and annotation data from one frame. Periodically "one-liner" listings are generated from the stored ID data. Each line of the one-liner listings contains 13 commonly referenced bits of information for one ERTS frame. The one-liner listings are distributed to ERTS Principal Investigators and others at LARS.

8.22 Actual Data Flow

The data flow procedure was implemented and operated as planned with exception of two steps. With respect to System Corrected Image (SYCI) data CCT's, no NASA Data Requests were required as the SYCI CCT products were received at LARS on a standing order basis just as 70 millimeter negative imagery products were received. This procedure for receiving SYCI CCT's, although unplanned, did expedite analysis projects since in many cases the CCT's were available to the LARS analyst as soon as he evaluated the photographic imagery and decided to use a particular CCT data set. The procedure did however have a disadvantage in that some ERTS frames received, did not find immediate application to the project. All ERTS frames received, however, are being filed in the LARSYS Multispectral Image Storage Tape Library.

The second exception to the operation of the data handling plan involved the on-liner data catalog computer-printed listing. The listing was to include a one line description of each ERTS frame for which CCT's were received. Due to the large number of frames received and subsequent storage problems, we were unable to maintain an up-to-date listing. The LARS ERTS Data Library Catalog file served in place of the on-liner listing adequately. For each entry, the catalog file includes frame ID, LARS Run Number, 70mm contact print of bands 5 and 7, and an indicator for receipt of the CCT's for the frame.

8.23 Reformatting Procedure

In order for LARS researchers to use the digital magnetic tape data representing ERTS multispectral scanner observations, the tapes received from NASA Data Processing Facility are reformatted or converted to the LARSYS Multispectral Image Storage Tape Format. The conversion is achieved by the use of computer programs especially developed to convert ERTS MSS computer compatible tapes to the LARSYS format.

The first step of the reformatting processing is receipt of an ERTS data reformatting request form. This form is completed by the LARS researcher requiring the data and is forwarded to the Reformatting Group. In general, an entire ERTS frame is not needed by the researcher, but rather a subframe or portion of the full frame. The reformatting request form allows the researcher to request the subframe area of a given ERTS frame in a number of ways. He may specify the upper, lower, left or right half of the frame; the test area by miles from the edges of the full frame, or the test area may be specified by latitude and longitude coordinates.

In general, the ERTS-to-LARSYS reformatting programs applies no corrections to the data radiometric values and does not alter the geometry of the data in any way. The program steps are to: Generate the required LARSYS tape header or identification record, calculate the sample set requested of the full ERTS frame, request the ERTS CCT's required and the LARSYS output data tape, and reformat line-by-line from the ERTS CCT's to the LARSYS tape. The LARSYS tape header is made up of information from the ERTS CCT's and reformatting program control data cards. The rectangular sample set to include the researchers requested test area is calculated in terms of frame scan line samples and line numbers. The calculation is based on the test area specification on the

researcher data reformatting request form and the frame center/spacecraft heading of the frame annotation record. When the CCT's from which scan line samples are requested are ready for reading and the LARSYS tape ready for writing, the program reads the required data, rearranges the samples into LARSYS format and writes the reformatted data, one scan line at a time.

After reformatting is completed the researcher requesting the data is notified via transmittal of a "Data Reformatting Notice" form. Receipt of this form notifies the researcher that the indicated data is available for analysis.

8.24 Reformatting Software

Eight significant data reformatting computer programs were written for the ERTS-A program contract. Several other small and insignificant programs were written for one-time use and special problem cases; these will not be specifically reported. In the following paragraphs, a general description of the significant programs is given.

MSS SYCI Reformatting

The program REFERTS was written to convert ERTS System Corrected Image computer compatible tapes to LARSYS-3 data tape format. Based on the user data reformatting request REFERTS reads one, two, three, or all four of the SYCI MSS CCT's, reformats the data into LARSYS-3 format, and writes the LARSYS data tape. The program performs no radiometric or geometric transformation. REFERTS can optionally reformat full frames or any portions of frames based on user requests. Subframes or portions may be specified by line and sample coordinates, miles from top and left frame edge coordinates, or longitude/latitude coordinates. Program output includes LARSYS-3 tape, data run descriptor form (see Figure 8.2) which is catalogued and distributed to users, and punched card run identification for entry into the LARSYS data

DATA STORAGE TAPE FILE

| | | | | |
|---------------------------|-----------------|--------------------------------|---------------------|------|
| RUN NUMBER..... | 72070600 | FLIGHTLINE ID..... | 106016491 | S.D. |
| DATE TAPE GENERATED... | MAR 1, 1974 | DATE DATA TAKEN..... | 9/21/72 | |
| TAPE NUMBER..... | 1126 | FILE... | 2 | |
| LINES OF DATA..... | 2340 | TIME DATA TAKEN..... | 0949 (LST) | |
| SECONDS OF DATA..... | 28.65 SEC | PLATFORM ALTITUDE..... | 3062000 | |
| AREA E-W 74 NM | N-S 100 NM | GROUND HEADING..... | 192 DEGREES | |
| LINE RATE..... | 81.68 LINES/SEC | FIELD OF VIEW | 8.59 DEG 0.1499 RAD | |
| TIME DATA WAS TAKEN..... | 1649 (GMT) | DATA SAMPLES/LINE/CHANNEL..... | 2428 | |
| SUN ELEVATION..... | 41 DEGREES | SAMPLE RATE | 0.0617 MILLIRADIANS | |
| SUN AZIMUTH..... | 147 DEGREES | LAT. AT FRAME CENTER..... | 44 D 30'N | |
| REVOLUTION NUMBER..... | 0836 | LONG. AT FRAME CENTER... | 097 D 45'W | |
| DAY SINCE LAUNCH..... | 060 | LAT. AT NADIR..... | 44 D 30'N | |
| SCENE/FRAME ID..... | 1060-1649100 | LONG. AT NADIR..... | 097 D 36'W | |
| FRAME ID..... | 3C100000 | RUN CENTER..... | 97D 28'W/ 44D 27'N | |
| STRIP ID..... | 0000 | AQUISITION SITE..... | GOLDSTONE | |
| SUN CALIBRATION DATA..... | | HI GAIN BAND 2..... | | |
| HI GAIN BAND 1..... | | RECORDED DATA..... | | |
| LINE LENGTH ADJUST..... | * | COMPRESSED DATA..... | * | |
| DIRECT DATA..... | * | DECOMPRESSION..... | * | |
| CALIBRATION WEDGE..... | | CALIBRATION..... | * | |

SPECTRAL BAND LIMITS IN MICROMETERS

| <u>CHAN</u> | <u>LOWER</u> | <u>UPPER</u> | <u>CHAN</u> | <u>LOWER</u> | <u>UPPER</u> | <u>CHAN</u> | <u>LOWER</u> | <u>UPPER</u> |
|-------------|--------------|--------------|-------------|--------------|--------------|-------------|--------------|--------------|
| (1) | 0.50 | 0.60 | (2) | 0.60 | 0.70 | (3) | 0.70 | 0.80 |
| (4) | 0.80 | 1.10 | (5) | ---- | ---- | (6) | ---- | ---- |
| (7) | ---- | ---- | (8) | ---- | ---- | (9) | ---- | ---- |
| (10) | ---- | ---- | (11) | ---- | ---- | (12) | ---- | ---- |

RUN CONDITIONS AND COMMENTS-- LINES 1 - 2340/1. COLUMNS 811 - 3232/1.

tape library cross-reference data set RUNTABLE.

Detector Statistics Processor

The program STATS was written to compute and display statistics for each of the 24 MSS processed detector outputs. Statistics calculated for each detector include variance, standard deviation and mean. In addition, combined statistics are computed for each group of detectors for each MSS band. Program input is the LARSYS-3 data tape. Program output includes histogram plots for each detector and band and printed tables of the computed statistics. Example program output is given in Figure 8.3.

Data Line Correction

The program FIX was written to provide a means of correcting "bad scan lines", that is, lines miscalibrated or lines for which the data is missing. The program reads a LARSYS-3 data tape and corrects specified bad lines, in one of two ways; (1) Replaces bad line N with line N-1, (2) Replaces bad line N with the average of lines N-1 and N+1. The correction method is specified on control cards as are the line numbers to be corrected. The program does not automatically detect bad lines.

Detector Calibration

The program SUBRUN has two basic functions: (1) applies a multiplicative and/or additive correction coefficient to each of the 24 MSS detector processed outputs and (2) selects a specified portion of the LARSYS-3 input data run for processing and output. The calibration transformation has the form

$$Y_{ij} = A_j X_{ij} + B_j$$

for i^{th} sample of detector j

j = 1, 24

i = 1, 1260480 for a full frame (2340 lines, 3232 samples)

STATISTICS FOR RUN 73001001, TAPE 1141, FILE 1

| CHANNEL 1 | DETECTOR | SAMPLES | SUM X**2 | SUM X | VARIANCE | STD. | DEV. | MEAN |
|-----------|----------|---------|--------------|--------------|----------|-------|------|--------|
| | 1 | 12090 | 0.66153E 07 | 0.267123E 06 | 59.109 | 7.688 | | 22.095 |
| | 2 | 12090 | 0.658144E 07 | 0.266372E 06 | 58.948 | 7.678 | | 22.032 |
| | 3 | 12090 | 0.650335E 07 | 0.265327E 06 | 58.539 | 7.506 | | 21.946 |
| | 4 | 12090 | 0.651527E 07 | 0.264618E 06 | 59.847 | 7.736 | | 21.887 |
| | 5 | 12090 | 0.635596E 07 | 0.262378E 06 | 54.746 | 7.399 | | 21.702 |
| | 6 | 12090 | 0.589895E 07 | 0.251517E 06 | 55.129 | 7.425 | | 20.804 |
| | TOTAL | 72540 | 0.384717E 08 | 0.157734E 07 | 57.535 | 7.585 | | 21.744 |
| CHANNEL 2 | DETECTOR | SAMPLES | SUM X**2 | SUM X | VARIANCE | STD. | DEV. | MEAN |
| | 1 | 12090 | 0.570863E 07 | 0.240906E 06 | 75.136 | 8.668 | | 19.926 |
| | 2 | 12090 | 0.545624E 07 | 0.236738E 06 | 67.880 | 8.239 | | 19.581 |
| | 3 | 12090 | 0.555454E 07 | 0.237976E 06 | 71.990 | 8.485 | | 19.684 |
| | 4 | 12090 | 0.546502E 07 | 0.236554E 06 | 69.202 | 8.319 | | 19.566 |
| | 5 | 12090 | 0.527388E 07 | 0.233125E 06 | 64.410 | 8.026 | | 19.282 |
| | 6 | 12090 | 0.515481E 07 | 0.230432E 06 | 63.102 | 7.944 | | 19.060 |
| | TOTAL | 72540 | 0.326127E 08 | 0.141573E 07 | 68.688 | 8.288 | | 19.517 |
| CHANNEL 3 | DETECTOR | SAMPLES | SUM X**2 | SUM X | VARIANCE | STD. | DEV. | MEAN |
| | 1 | 12090 | 0.559810E 07 | 0.245282E 06 | 51.437 | 7.172 | | 20.288 |
| | 2 | 12090 | 0.533132E 07 | 0.238855E 06 | 50.639 | 7.116 | | 19.756 |
| | 3 | 12090 | 0.574626E 07 | 0.248340E 06 | 53.364 | 7.305 | | 20.541 |
| | 4 | 12090 | 0.518346E 07 | 0.233877E 06 | 54.527 | 7.384 | | 19.345 |
| | 5 | 12090 | 0.536764E 07 | 0.238287E 06 | 55.516 | 7.451 | | 19.709 |
| | 6 | 12090 | 0.549242E 07 | 0.241013E 06 | 56.898 | 7.543 | | 19.935 |
| | TOTAL | 72540 | 0.327185E 08 | 0.144565E 07 | 53.876 | 7.340 | | 19.929 |
| CHANNEL 4 | DETECTOR | SAMPLES | SUM X**2 | SUM X | VARIANCE | STD. | DEV. | MEAN |
| | 1 | 12090 | 0.127571E 07 | 0.117870E 06 | 10.456 | 3.234 | | 9.750 |
| | 2 | 12090 | 0.127149E 07 | 0.117540E 06 | 10.651 | 3.264 | | 9.722 |
| | 3 | 12090 | 0.119045E 07 | 0.112680E 06 | 11.593 | 3.405 | | 9.321 |
| | 4 | 12090 | 0.116506E 07 | 0.111664E 06 | 11.311 | 3.363 | | 9.236 |
| | 5 | 12090 | 0.132350E 07 | 0.119386E 06 | 11.712 | 3.422 | | 9.815 |
| | 6 | 12090 | 0.118477E 07 | 0.112581E 06 | 11.285 | 3.359 | | 9.312 |
| | TOTAL | 72540 | 0.741099E 07 | 0.691735E 06 | 11.231 | 3.351 | | 9.536 |

Figure 8.3a Example Output for Detector Statistics Processor.

HISTOGRAM FOR RUN 7301C01 , CHANNEL 4

SAMPLES = 72540. SUM OF X**2 = 0.741098 J7 SUM OF X = 0.691735E 06
 VARIANCE = 11.231 STD.DEV. = 3.351 MEAN = 9.536

EACH * = 291 SAMPLES.

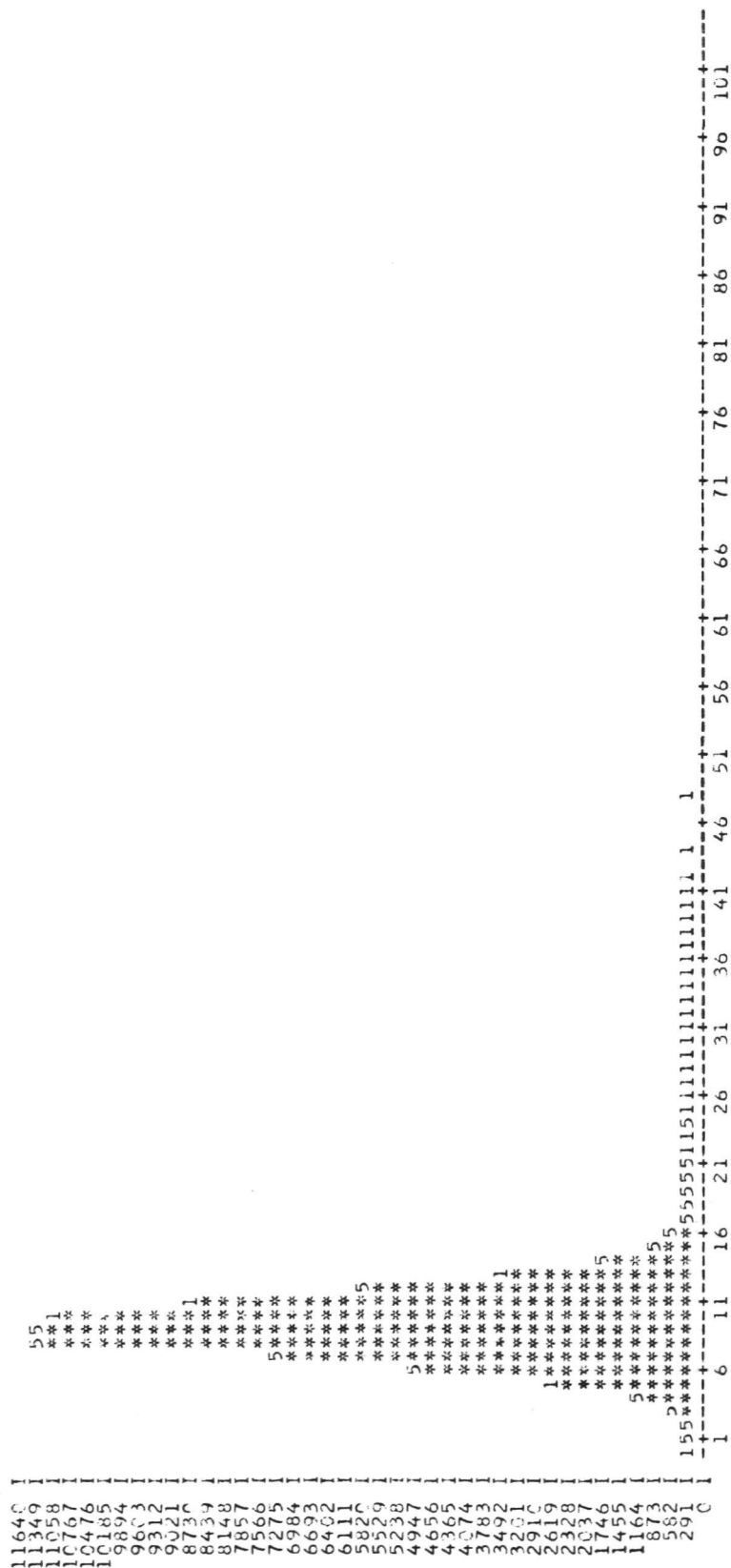


Figure 8.3b Example Histogram Output for Detector Statistics Processor.

The 24 A and 24 B coefficients are computed by the program STATS discussed above.

Connection of Frames

The program CONECT was written in order to facilitate automated joining of adjacent frames of a given orbit pass. An automatic and specialized program was required since contiguous MSS CCT data frames within an orbit were found to be overlapping by an inconsistent number of scan lines (approximately 300). The method used to locate a connecting point is as follows: (1) Read and store three segments of the first scan line of the second or Southern-most of the frames to be connected, (2) beginning at line 1900 of the Northern frame, compare line-by-line, like segments with the three stored segments, (3) when segments of a line of the Northern frame compare with a 60 or more percent equality, the connection point is located, (4) the output run is then generated by writing lines from the Northern frame to and including the connect point line, followed by writing lines of the Southern frame from line 2. The 60 percent equality criterion was derived by experience after noting that overlapping lines of two frames were not generally identical.

One-Line Listing of ERTS Frames

The program ONELIN was written to provide a computerized catalogue listing of all CCT data frames received. The program performs the functions of reading and storing header and annotation information from the ERTS CCT's and printing listings of the stored information. One line of data is printed for each frame and the listing may be sorted by scene frame ID or LARSYS Run Number. Example output is shown in Figure 8.4.

Geometric Transformation

The program GEMCOR was written to geometrically transform ERTS SYCI MSS data by predetermined parameters. Program input is

BULK ERTS DATA
TAPE INFORMATION

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

JUNE 20, 1974
6 22 12 PM

| U P | LARS RUN NUMBER | SCENE/FRAME ID | DATE DATA TAKEN | TIME DATA TAKEN | REVOL. NUMBER | FRAME CENTER | SUN ELEVATION AZIMUTH | FIRST BULK TAPE | DATE TAPES RECEIVED | A Q U | CC | USER ID |
|--------|-----------------------|-------------------|-----------------------|-----------------------|------------------|------------------|-----------------------------|-----------------------|---------------------------|-------------|----|------------|
| 7 | 720326 | 1003-1633400 | 7/26/72 | 1633 | 0041 | 094-50 W/41-26 N | 57/124 | 1021 | 10/12/72 | N | 27 | UN 127 |
| 2 | 720327 | 1017-1609100 | 8/09/72 | 1609 | 0236 | 087-54 W/43-54 N | 53/132 | 1025 | 10/12/72 | G | 27 | * 101 |
| 7 | 720328 | 1017-1609300 | 8/09/72 | 1609 | 0236 | 088-26 W/42-29 N | 54/130 | 1029 | 10/12/72 | G | 27 | * 101 |
| 5 | 720329 | 1033-1558200 | 8/25/72 | 1558 | 0459 | 085-36 W/41-39 N | 51/135 | 1033 | 10/12/72 | N | 27 | * 101 |
| 6 | 720330 | 1029-1719500 | 8/21/72 | 1719 | 0404 | 107-23 W/37-49 N | 54/128 | 1037 | 10/16/72 | G | 27 | UN 103 |
| 7 | 720334 | 1081-1705501 | 10/12/72 | 1705 | 1129 | 100-50 W/47-24 N | 31/155 | 1045 | 11/7/72 | G | 27 | UN 630 |
| 7 | 720335 | 1057-1632300 | 9/18/72 | 1632 | 0794 | 093-59 W/43-02 N | 43/145 | 1049 | 11/7/72 | N | 27 | UN 630 |
| 9 | 720337 | 1044-1659300 | 9/05/72 | 1659 | 0613 | 098-54 W/48-45 N | 43/146 | 1057 | 11/13/72 | G | 27 | UN 630 |
| 9 | 720338 | 1052-1605200 | 9/13/72 | 1605 | 0724 | 088-45 W/37-24 N | 48/138 | 1061 | 11/13/72 | N | 27 | UN 127 |
| 0 | 720339 | 1052-1605000 | 9/13/72 | 1605 | 0724 | 088-17 W/38-50 N | 47/139 | 1065 | 11/13/72 | N | 27 | UN 127 |
| 2 | 720340 | 1017-1610200 | 8/09/72 | 1610 | 0236 | 089-27 W/39-39 N | 55/126 | 1069 | 11/13/72 | G | 27 | UN 127 |
| 2 | 720341 | 1017-1610000 | 8/09/72 | 1610 | 0236 | 088-57 W/41-04 N | 55/128 | 1073 | 11/13/72 | G | 27 | UN 127 |
| 4 | 720342 | 1034-1605200 | 8/26/72 | 1605 | 0473 | 088-49 W/37-22 N | 53/130 | 1077 | 11/13/72 | N | 27 | UN 127 |
| 4 | 720343 | 1050-1553500 | 9/11/72 | 1553 | 0696 | 085-53 W/37-26 N | 49/137 | 1081 | 11/13/72 | N | 27 | UN 127 |
| 6 | 720346 | 1086-1554101 | 10/11/72 | 1554 | 1198 | 085-56 W/37-26 N | 38/149 | 1093 | 11/15/72 | N | 27 | UN 127 |
| 6 | 720348 | 1080-1703501 | 10/11/72 | 1703 | 1115 | 104-01 W/34-37 N | 42/146 | 1105 | 11/22/72 | G | 27 | UN 630 |
| 7 | 720349 | 1088-1605501 | 10/19/72 | 1605 | 1226 | 088-48 W/37-24 N | 37/150 | 1109 | 11/20/72 | N | 27 | UN 127 |
| 7 | 720350 | 1088-1605201 | 10/19/72 | 1605 | 1226 | 088-20 W/38-50 N | 36/151 | 1113 | 11/20/72 | N | 27 | UN 127 |
| 9 | 720351 | 1044-1659500 | 9/05/72 | 1659 | 0613 | 099-31 W/47-20 N | 44/145 | 1117 | 11/20/72 | G | 27 | UN 630 |
| 0 | 720354 | 1088-1604301 | 10/19/72 | 1604 | 1226 | 087-20 W/41-40 N | 34/153 | 1129 | 11/20/72 | N | 27 | UN 127 |
| 1 | 720355 | 1079-1658301 | 10/10/72 | 1658 | 1101 | 103-02 W/33-12 N | 43/144 | 1133 | 11/22/72 | G | 27 | UN 630 |
| 2 | 720358 | 1016-1605500 | 8/08/72 | 1605 | 0222 | 089-20 W/35-39 N | 57/119 | 1097 | 11/20/72 | N | 27 | UN 127 |
| 3 | 720359 | 1051-1559400 | 9/12/72 | 1559 | 0710 | 087-21 W/37-27 N | 48/137 | 1153 | 11/22/72 | N | 27 | UN 127 |
| 4 | 720362 | 1080-1704401 | 10/11/72 | 1704 | 1115 | 104-54 W/31-46 N | 44/143 | 1165 | 11/22/72 | G | 27 | UN 630 |
| 5 | 720363 | 1079-1658001 | 10/10/72 | 1658 | 1101 | 102-36 W/34-37 N | 42/145 | 1169 | 11/22/72 | G | 27 | UN 630 |
| 6 | 720364 | 1003-1817500 | 7/26/72 | 1817 | 0042 | 121-46 W/38-06 N | 58/118 | 1173 | 11/22/72 | G | 27 | I 84 |
| 7 | 720365 | 1021-1632400 | 8/13/72 | 1632 | 0292 | 094-19 W/42-08 N | 53/131 | 1177 | 11/22/72 | G | 27 | UN 630 |
| 8 | 720366 | 1079-1658501 | 10/10/72 | 1658 | 1101 | 103-28 W/31-46 N | 44/143 | 1181 | 11/22/72 | G | 27 | UN 630 |
| 9 | 720368 | 1078-1653101 | 10/09/72 | 1653 | 1087 | 102-01 W/31-47 N | 44/143 | 1201 | 11/27/72 | G | 27 | UN 630 |
| 0 | 720369 | 1078-1652401 | 10/09/72 | 1652 | 1087 | 101-35 W/33-13 N | 43/144 | 1205 | 11/27/72 | G | 27 | UN 630 |
| 1 | 720370 | 1080-1704101 | 10/11/72 | 1704 | 1115 | 104-28 W/33-12 N | 43/145 | 1101 | 11/20/72 | G | 27 | UN 630 |
| 2 | 720371 | 1077-1647201 | 10/08/72 | 1647 | 1073 | 100-33 W/31-47 N | 45/142 | 1213 | 11/27/72 | G | 27 | UN 630 |
| 3 | 720376 | 1077-1647001 | 10/08/72 | 1647 | 1073 | 100-07 W/33-12 N | 44/143 | 1233 | 11/27/72 | G | 27 | UN 630 |
| 4 | 720381 | 1101-1720301 | 11/01/72 | 1720 | 1408 | 107-31 W/37-20 N | 33/152 | 1253 | 11/27/72 | G | 27 | UN 103 |
| 5 | 720382 | 1100-1714001 | 10/31/72 | 1714 | 1394 | 105-06 W/40-11 N | 31/154 | 1257 | 11/27/72 | G | 27 | UN 103 |
| 6 | 720383 | 1103-1548501 | 11/03/72 | 1548 | 1435 | 084-36 W/37-16 N | 33/153 | 1261 | 11/27/72 | N | 27 | UN 127 |
| 7 | 720384 | 1086-1553501 | 10/17/72 | 1553 | 1198 | 085-28 W/38-51 N | 37/150 | 1265 | 11/27/72 | N | 27 | UN 630 |
| 8 | 720385 | 1086-1553201 | 10/17/72 | 1553 | 1198 | 084-59 W/40-16 N | 36/151 | 1273 | 11/27/72 | N | 27 | UN 127 |
| 9 | 720386 | 1089-1610401 | 10/20/72 | 1610 | 1240 | 089-16 W/40-16 N | 35/152 | 1277 | 11/27/72 | N | 27 | UN 127 |
| 0 | 720387 | 1089-1611101 | 10/20/72 | 1611 | 1240 | 089-46 W/38-50 N | 36/151 | 1281 | 11/29/72 | N | 27 | UN 127 |
| 1 | 720388 | 1089-1610201 | 10/20/72 | 1610 | 1240 | 088-46 W/41-41 N | 33/153 | 1285 | 11/29/72 | N | 27 | UN 127 |
| 2 | 720392 | 1105-1609201 | 11/05/72 | 1600 | 1463 | 087-27 W/37-16 N | 32/153 | 1309 | 12/11/72 | N | 27 | UN 127 |
| 3 | 720393 | 1053-1755400 | 9/14/72 | 1755 | 0739 | 117-24 W/33-07 N | 50/133 | 1313 | 12/11/72 | G | 27 | * 1052 |
| 4 | 720395 | 1097-1658201 | 10/28/72 | 1658 | 1352 | 102-42 W/34-28 N | 37/150 | 1321 | 12/11/72 | G | 27 | UN 630 |
| 5 | 720411 | 1098-1704101 | 10/29/72 | 1704 | 1366 | 104-07 W/34-31 N | 36/150 | 1325 | 12/14/72 | G | 27 | UN 630 |
| 6 | 720412 | 1105-1559501 | 11/05/72 | 1559 | 1463 | 086-58 W/38-41 N | 31/154 | 1329 | 12/14/72 | N | 27 | UN 630 |
| 7 | 720413 | 1106-1603401 | 11/06/72 | 1603 | 1477 | 088-26 W/38-42 N | 31/154 | 1337 | 12/14/72 | N | 27 | UN 127 |
| 8 | 720414 | 1106-1606001 | 11/06/72 | 1606 | 1477 | 088-54 W/37-17 N | 32/153 | 1341 | 12/14/72 | N | 27 | UN 127 |
| 9 | 720416 | 1098-1704301 | 10/29/72 | 1704 | 1366 | 104-34 W/33-05 N | 37/149 | 1353 | 12/14/72 | G | 27 | UN 630 |
| 0 | 720417 | 1040-1454300 | 9/21/72 | 1454 | 0556 | 069-35 W/44-33 N | 47/141 | 1357 | 12/14/72 | N | 27 | * 57 |
| 1 | 720418 | 1062-1518100 | 9/23/72 | 1518 | 0863 | 075-49 W/41-40 N | 42/146 | 1361 | 12/14/72 | N | 27 | * 57 |
| 2 | 720419 | 1080-1519201 | 10/11/72 | 1519 | 1114 | 076-48 W/38-54 N | 38/149 | 1365 | 12/14/72 | N | 27 | * 101 |
| 3 | 720423 | 1114-1652501 | 11/14/72 | 1652 | 1589 | 101-15 W/34-34 N | 32/152 | 1373 | 12/14/72 | G | 27 | UN 630 |
| 4 | 720425 | 1079-1511501 | 10/10/72 | 1511 | 1100 | 073-18 W/44-35 N | 34/153 | 1381 | 12/14/72 | N | 27 | * 57 |
| 5 | 720426 | 1063-1525400 | 9/24/72 | 1525 | 0877 | 079-10 W/35-57 N | 46/141 | 1385 | 12/14/72 | N | 27 | * 57 |
| 6 | 720427 | 1057-1634300 | 9/18/72 | 1634 | 0794 | 096-26 W/35-54 N | 48/138 | 1389 | 12/14/72 | N | 27 | * 57 |
| 7 | 720428 | 1113-1647101 | 11/13/72 | 1647 | 1575 | 099-51 W/34-25 N | 32/152 | 1125 | 12/20/72 | G | 27 | UN 630 |
| 8 | 720429 | 1113-1648001 | 11/13/72 | 1648 | 1575 | 100-43 W/31-33 N | 34/150 | 1137 | 12/20/72 | G | 27 | UN 630 |
| 9 | 720430 | 1113-1647301 | 11/13/72 | 1647 | 1575 | 100-17 W/32-59 N | 33/151 | 1185 | 12/20/72 | G | 27 | UN 630 |
| 0 | 720432 | 1116-1704201 | 11/16/72 | 1704 | 1617 | 104-06 W/34-36 N | 31/152 | 1193 | 12/20/72 | G | 27 | UN 630 |
| 1 | 720433 | 1116-1704501 | 11/16/72 | 1704 | 1617 | 104-32 W/33-10 N | 32/152 | 1197 | 12/20/72 | G | 27 | UN 630 |
| 2 | 720434 | 1045-1524300 | 9/06/72 | 1524 | 0626 | 077-20 W/40-12 N | 48/138 | 1209 | 12/20/72 | N | 27 | * 57 |
| 3 | 720435 | 1029-1536100 | 8/21/72 | 1536 | 0403 | 081-00 W/39-25 N | 53/130 | 1269 | 12/20/72 | N | 27 | * 57 |
| 4 | 720436 | 1010-0837500 | 8/02/72 | 0837 | 0134 | 023-04 E/37-20 N | 57/119 | 1397 | 12/20/72 | A | 27 | F 375 |
| 5 | 720437 | 1010-0837200 | 8/02/72 | 0837 | 0134 | 023-30 E/38-40 N | 57/122 | 1393 | 12/20/72 | A | 27 | F 375 |
| 6 | 720440 | 1119-1720401 | 11/19/72 | 1720 | 1659 | 107-30 W/37-25 N | 28/154 | 1333 | 12/21/72 | G | 27 | UN 103 |
| 7 | 720441 | 1116-1705101 | 11/16/72 | 1705 | 1617 | 104-57 W/31-43 N | 34/151 | 1349 | 12/21/72 | G | 27 | UN 630 |
| 8 | 720445 | 1115-1659501 | 11/15/72 | 1659 | 1603 | 103-57 W/30-16 N | 35/150 | 1409 | 12/27/72 | G | 27 | UN 630 |
| 9 | 720447 | 1036-1616500 | 8/28/72 | 1616 | 0501 | 091-38 W/37-37 N | 52/131 | 1413 | 12/27/72 | G | 27 | * 57 |
| 0 | 720448 | 1045-1525400 | 9/06/72 | 1525 | 0626 | 079-15 W/35-06 N | 51/133 | 1417 | 12/27/72 | N | 27 | * 57 |
| 1 | 720450 | 1114-1653201 | 11/14/72 | 1653 | 1589 | 101-41 W/33-08 N | 33/151 | 1425 | 12/27/72 | G | 27 | UN 630 |
| 2 | 720453 | 1113-1559101 | 11/23/72 | 1559 | 1714 | 085-56 W/41-39 N | 24/156 | 1465 | 1/8/73 | N | 27 | UN 127 |

Figure 8.4a One Line Listing of ERTS CCT Data in File by LARS Run Number.

| D U P | LARS RUN NUMBER | SCENE/FRAME ID | DATE DATA TAKEN | TIME DATA TAKEN | REVOL. NUMBER | FRAME CENTER | SUN ELEVATION AZIMUTH | FIRST BULK TAPE | DATE TAPES RECEIVED | A Q U | CC | USER ID |
|-------------|-----------------------|-------------------|-----------------------|-----------------------|------------------|------------------|-----------------------------|-----------------------|---------------------------|-------------|----|------------|
| 1 | 720326 | 1003-1633400 | 7/26/72 | 1633 | 0041 | 094-50 W/41-26 N | 57/124 | 1021 | 10/12/72 | N | 27 | UN 127 |
| 2 | 720364 | 1003-1817500 | 7/26/72 | 1817 | 0042 | 121-46 W/38-06 N | 58/118 | 1173 | 11/22/72 | N | 27 | I 84 |
| 3 | 720645 | 1006-1652200 | 7/29/72 | 1652 | 0083 | 101-35 W/33-06 N | 59/110 | 2397 | 4/ 4/73 | N | 27 | UN 630 |
| 4 | 720642 | 1007-1655400 | 7/30/72 | 1655 | 0097 | 100-14 W/41-38 N | 56/125 | 2385 | 4/ 4/73 | N | 27 | UN 630 |
| 5 | 720643 | 1007-1656300 | 7/30/72 | 1656 | 0097 | 101-14 W/38-46 N | 57/121 | 2389 | 4/ 4/73 | N | 27 | UN 630 |
| 6 | 720653 | 1007-1657500 | 7/30/72 | 1657 | 0097 | 102-41 W/34-17 N | 59/113 | 1245 | 4/17/73 | G | 27 | UN 630 |
| 7 | 720437 | 1010-0837200 | 8/02/72 | 0837 | 0134 | 023-32 E/38-46 N | 57/122 | 1393 | 12/20/72 | A | 27 | F 375 |
| 8 | 720436 | 1010-0837500 | 8/02/72 | 0837 | 0134 | 023-04 E/37-20 N | 57/119 | 1397 | 12/20/72 | A | 27 | F 375 |
| 9 | 720693 | 1010-1403300 | 8/02/72 | 1403 | 0137 | 068-16 W/17-34 S | 38/049 | 3357 | 9/17/73 | A | 27 | UN 127 |
| 10 | 720672 | 1013-1729400 | 8/05/72 | 1729 | 0181 | 108-17 W/43-10 N | 54/130 | 2473 | 7/ 2/73 | G | 27 | * 321 |
| 11 | 720673 | 1014-1735200 | 8/06/72 | 1735 | 0195 | 109-43 W/43-09 N | 54/130 | 2477 | 7/ 2/73 | G | 27 | * 321 |
| 12 | 720358 | 1016-1605500 | 8/08/72 | 1605 | 0222 | 089-20 W/35-39 N | 57/119 | 1097 | 11/20/72 | N | 27 | UN 127 |
| 13 | 720685 | 1017-1105500 | 8/09/72 | 1105 | 0233 | 017-16 W/21-37 N | 59/095 | 9999 | 8/13/73 | A | 27 | * 105 |
| 14 | 720327 | 1017-1609100 | 8/09/72 | 1609 | 0236 | 087-54 W/43-54 N | 53/132 | 1025 | 10/12/72 | G | 27 | * 101 |
| 15 | 720328 | 1017-1609300 | 8/09/72 | 1609 | 0236 | 088-26 W/42-29 N | 54/130 | 1029 | 10/12/72 | G | 27 | * 101 |
| 16 | 720341 | 1017-1610000 | 8/09/72 | 1610 | 0236 | 088-57 W/41-04 N | 55/128 | 1073 | 11/13/72 | G | 27 | UN 127 |
| 17 | 720340 | 1017-1610200 | 8/09/72 | 1610 | 0236 | 089-27 W/39-39 N | 55/126 | 1069 | 11/13/72 | G | 27 | UN 127 |
| 18 | 720652 | 1019-1622000 | 8/11/72 | 1622 | 0264 | 092-25 W/39-22 N | 55/126 | 1241 | 4/17/73 | N | 27 | UA 73 |
| 19 | 720655 | 1020-1627500 | 8/12/72 | 1627 | 0278 | 094-03 W/38-46 N | 55/126 | 1289 | 4/23/73 | G | 27 | UA 73 |
| 20 | 720365 | 1021-1632400 | 8/13/72 | 1632 | 0292 | 094-19 W/42-08 N | 53/131 | 1177 | 11/22/72 | G | 27 | UN 630 |
| 21 | 720620 | 1021-1633300 | 8/13/72 | 1633 | 0292 | 095-19 W/39-17 N | 55/127 | 2189 | 3/19/73 | G | 27 | UA 72 |
| 22 | 720649 | 1021-1634200 | 8/13/72 | 1634 | 0292 | 096-16 W/36-26 N | 56/122 | 1225 | 4/17/73 | G | 27 | UA 73 |
| 23 | 720524 | 1023-1500300 | 8/15/72 | 1500 | 0319 | 070-58 W/43-29 N | 52/134 | 1529 | 1/15/73 | N | 27 | * 57 |
| 24 | 720651 | 1024-1652200 | 8/16/72 | 1652 | 0334 | 101-28 W/33-30 N | 56/119 | 1237 | 4/17/73 | G | 27 | UN 630 |
| 25 | 720607 | 1024-1652500 | 8/16/72 | 1652 | 0334 | 101-54 W/32-04 N | 57/117 | 2009 | 3/ 7/73 | G | 27 | UN 630 |
| 26 | 720650 | 1025-1656000 | 8/17/72 | 1656 | 0348 | 100-33 W/40-44 N | 53/130 | 1229 | 4/17/73 | N | 27 | UN 630 |
| 27 | 720656 | 1025-1658000 | 8/17/72 | 1658 | 0348 | 102-51 W/33-43 N | 56/120 | 1293 | 4/23/73 | G | 27 | UN 630 |
| 28 | 720635 | 1027-1523300 | 8/19/72 | 1523 | 0375 | 076-41 W/43-13 N | 51/135 | 2301 | 3/27/73 | N | 27 | UA 73 |
| 29 | 720682 | 1028-1530200 | 8/20/72 | 1530 | 0389 | 079-30 W/39-40 N | 53/130 | 2825 | 8/ 6/73 | N | 27 | UA 73 |
| 30 | 720435 | 1029-1536100 | 8/21/72 | 1536 | 0403 | 081-00 W/39-25 N | 53/130 | 1269 | 12/20/72 | N | 27 | * 57 |
| 31 | 720330 | 1029-1719500 | 8/21/72 | 1719 | 0404 | 107-23 W/37-49 N | 54/128 | 1037 | 10/16/72 | G | 27 | UN 103 |
| 32 | 720679 | 1029-1721300 | 8/21/72 | 1721 | 0404 | 109-11 W/32-05 N | 56/120 | 2805 | 8/ 1/73 | G | 27 | * 105 |
| 33 | 720548 | 1033-1558000 | 8/25/72 | 1558 | 0459 | 085-24 W/43-34 N | 50/137 | 1625 | 1/22/73 | N | 27 | UA 72 |
| 34 | 720329 | 1033-1558200 | 8/25/72 | 1558 | 0459 | 085-56 W/41-39 N | 51/135 | 1033 | 10/12/72 | N | 27 | * 101 |
| 35 | 720644 | 1033-1559100 | 8/25/72 | 1559 | 0459 | 000-00 W/38-48 N | 52/131 | 2393 | 4/ 4/73 | N | 27 | UN 630 |
| 36 | 720523 | 1034-1605000 | 8/26/72 | 1605 | 0473 | 088-21 W/38-46 N | 52/132 | 1525 | 1/15/73 | N | 27 | * 57 |
| 37 | 720342 | 1034-1605200 | 8/26/72 | 1605 | 0473 | 088-49 W/37-22 N | 53/130 | 1077 | 11/13/72 | N | 27 | UN 127 |
| 38 | 720630 | 1035-1611200 | 8/27/72 | 1611 | 0487 | 090-30 W/36-34 N | 53/129 | 2253 | 3/26/73 | N | 27 | UN 127 |
| 39 | 720447 | 1036-1616500 | 8/28/72 | 1616 | 0501 | 091-38 W/37-37 N | 52/131 | 1413 | 12/27/72 | G | 27 | * 57 |
| 40 | 720633 | 1037-1622100 | 8/29/72 | 1622 | 0515 | 092-40 W/38-50 N | 51/133 | 2269 | 3/26/73 | G | 27 | UA 73 |
| 41 | 720544 | 1038-1630300 | 8/30/72 | 1630 | 0529 | 096-44 W/30-16 N | 55/122 | 1609 | 1/22/73 | G | 27 | UA 73 |
| 42 | 720417 | 1040-1454300 | 9/01/72 | 1454 | 0556 | 069-05 W/44-33 N | 47/141 | 1357 | 12/14/72 | N | 27 | * 57 |
| 43 | 720640 | 1043-1655500 | 9/04/72 | 1655 | 0599 | 100-18 W/41-38 N | 48/139 | 2373 | 4/ 4/73 | N | 27 | UN 630 |
| 44 | 720531 | 1044-0827200 | 9/05/72 | 0827 | 0608 | 025-02 E/34-30 N | 52/130 | 1557 | 1/19/73 | A | 27 | F 375 |
| 45 | 720540 | 1044-1518000 | 9/05/72 | 1518 | 0612 | 075-28 W/42-46 N | 47/140 | 1593 | 1/19/73 | N | 27 | UA 72 |
| 46 | 720337 | 1044-1659300 | 9/05/72 | 1659 | 0613 | 098-54 W/48-45 N | 43/146 | 1057 | 11/13/72 | G | 27 | UN 630 |
| 47 | 720351 | 1044-1659500 | 9/05/72 | 1659 | 0613 | 099-31 W/47-20 N | 44/145 | 1117 | 11/20/72 | G | 27 | UN 630 |
| 48 | 720434 | 1045-1524300 | 9/06/72 | 1524 | 0626 | 077-50 W/40-12 N | 48/138 | 1209 | 12/20/72 | N | 27 | * 57 |
| 49 | 720448 | 1045-1525400 | 9/06/72 | 1525 | 0626 | 079-15 W/35-56 N | 51/133 | 1417 | 12/27/72 | N | 27 | * 57 |
| 50 | 720674 | 1046-1530100 | 9/07/72 | 1530 | 0640 | 079-15 W/40-10 N | 48/138 | 2629 | 7/23/73 | N | 27 | UA 73 |
| 51 | 720686 | 1046-1531500 | 9/07/72 | 1531 | 0640 | 081-07 W/34-27 N | 51/132 | 3005 | 8/20/73 | N | 27 | UN 103 |
| 52 | 720619 | 1047-1719100 | 9/08/72 | 1719 | 0655 | 106-28 W/40-22 N | 48/139 | 2185 | 3/19/73 | N | 27 | UN 103 |
| 53 | 720622 | 1047-1720000 | 9/08/72 | 1720 | 0655 | 107-26 W/37-30 N | 49/136 | 2201 | 3/22/73 | G | 27 | UN 103 |
| 54 | 720343 | 1050-1553500 | 9/11/72 | 1553 | 0696 | 085-53 W/37-26 N | 49/137 | 1081 | 11/13/72 | N | 27 | UN 127 |
| 55 | 720359 | 1051-1559400 | 9/12/72 | 1559 | 0710 | 087-21 W/37-27 N | 48/137 | 1153 | 11/22/72 | N | 27 | UN 127 |
| 56 | 720339 | 1052-1605000 | 9/13/72 | 1605 | 0724 | 088-17 W/38-50 N | 47/139 | 1065 | 11/13/72 | N | 27 | UN 127 |
| 57 | 720338 | 1052-1605200 | 9/13/72 | 1605 | 0724 | 088-45 W/37-24 N | 48/138 | 1061 | 11/13/72 | N | 27 | UN 127 |
| 58 | 720629 | 1052-1605500 | 9/13/72 | 1605 | 0724 | 089-13 W/35-50 N | 49/136 | 2249 | 3/26/73 | N | 27 | UN 127 |
| 59 | 720638 | 1052-1749500 | 9/13/72 | 1749 | 0725 | 115-58 W/33-07 N | 50/133 | 2353 | 4/ 3/73 | G | 27 | * 52 |
| 60 | 720573 | 1053-1609500 | 9/14/72 | 1609 | 0738 | 088-46 W/41-41 N | 45/142 | 1705 | 2/ 5/73 | N | 27 | UN 127 |
| 61 | 720393 | 1053-1755400 | 9/14/72 | 1755 | 0739 | 117-24 W/33-07 N | 50/133 | 1313 | 12/11/72 | G | 27 | * 1052 |
| 62 | 720335 | 1057-1632300 | 9/18/72 | 1632 | 0794 | 093-59 W/43-02 N | 43/145 | 1049 | 11/ 7/72 | N | 27 | UN 630 |
| 63 | 720427 | 1057-1634300 | 9/18/72 | 1634 | 0794 | 096-26 W/35-54 N | 48/138 | 1389 | 12/14/72 | N | 27 | * 57 |
| 64 | 720596 | 1060-1506000 | 9/21/72 | 1506 | 0835 | 071-56 W/44-30 N | 41/147 | 1869 | 2/26/73 | N | 27 | * 57 |
| 65 | 720646 | 1061-1655500 | 9/22/72 | 1655 | 0850 | 100-13 W/41-40 N | 43/145 | 1161 | 4/13/73 | G | 27 | UN 630 |
| 66 | 720639 | 1061-1656400 | 9/22/72 | 1656 | 0850 | 101-12 W/38-48 N | 44/143 | 2357 | 4/ 3/73 | G | 27 | UN 630 |
| 67 | 720418 | 1062-1518100 | 9/23/72 | 1518 | 0863 | 075-49 W/41-40 N | 42/146 | 1361 | 12/14/72 | N | 27 | * 57 |
| 68 | 720426 | 1063-1525400 | 9/24/72 | 1525 | 0877 | 079-10 W/35-57 N | 46/141 | 1385 | 12/14/72 | N | 27 | * 57 |
| 69 | 720675 | 1066-1725100 | 9/27/72 | 1725 | 0920 | 108-19 W/38-54 N | 43/145 | 2617 | 7/23/73 | G | 27 | UN 103 |
| 70 | 720647 | 1066-1725400 | 9/27/72 | 1725 | 0920 | 108-47 W/37-28 N | 44/143 | 1189 | 4/16/73 | G | 27 | UN 103 |
| 71 | 720535 | 1069-1558001 | 9/30/72 | 1558 | 0961 | 085-19 W/43-10 N | 39/149 | 1573 | 1/19/73 | N | 27 | UA 72 |
| 72 | 720536 | 1069-1558501 | 9/30/72 | 1558 | 0961 | 086-31 W/43-10 N | 41/147 | 1577 | 1/19/73 | N | 27 | UA 72 |

Figure 8.4b One Line Listing of ERTS CCT Data in File by Scene Frame ID.

the LARSYS data tape and control parameters. The program performs the following geometric alterations: Along track to cross track sampling aspect ratio, earth rotational skew, rotation for North orientation, aspect change for line printer gray mapping, and re-scaling to 1:24000. A detailed discussion of the geometric correction parameters are given in Section 8.5.

Scene Corrected Image Reformatting

The program PRECIS was written to convert ERTS MSS Scene Corrected Image CCT data tapes to LARSYS-3 format. The program is used to reformat a full frame or any portion of a frame.

8.25 Data Products

A summary of ERTS MSS data received and processed is shown in Table 8.1. ERTS frames included represent MSS System Corrected Image (SYCI) data with exception of eight MSS Scene Corrected Image (SCCI) data sets. Shown in Table 8.2 is a list of types of and numbers of LARSYS data runs generated during the ERTS contract period.

TABLE 8.1
Summary of ERTS CCT Frames
Received and Processed

| USER/ID | NUMBER FRAMES | NUMBER FRAMES | PERCENT USED |
|----------|------------------|------------------|-----------------|
| UNLISTED | 9 | 7 | 77.8 |
| U 127 | 215 | 78 | 36.3 |
| 101 Z | 7 | 5 | 71.4 |
| U 103 | 88 | 29 | 33.0 |
| U 630 | 724 | 73 | 10.1 |
| I 084 | 9 | 9 | 100.0 |
| 105 Z | 9 | 8 | 88.9 |
| U 057 | 20 | 16 | 80.0 |
| F 375 | 3 | 2 | 66.7 |
| N 374 | 7 | 3 | 42.9 |
| UA 72 | 13 | 12 | 92.3 |
| UA 73 | 85 | 73 | 85.9 |
| U 168 | 8 | 0 | 0.0 |
| S 351 | 3 | 0 | 0.0 |
| P 169 | 2 | 2 | 100.0 |
| A 328 | 8 | 8 | 100.0 |
| U 321 | 2 | 2 | 100.0 |
| 414 | 1 | 0 | 0.0 |
| 443 | 1 | 0 | 0.0 |
| 130 | 2 | 1 | 50.0 |
| UA 71 | 4 | 3 | 75.0 |
| 142 | 1 | 1 | 100.0 |
| 118 Z | 41 | 23 | 56.1 |
| 1040 AA | 12 | 1 | 8.3 |
| 1708 AA | 1 | 1 | 100.0 |
| 1049 AA | 55 | 0 | 0.0 |
| 1050 AA | 30 | 0 | 0.0 |
| 1159 AD | 9 | 6 | 66.7 |
| 100 A | 14 | 6 | 42.9 |
| 226 A | 1 | 1 | 100.0 |
| 1569 CA | 1 | 1 | 100.0 |
| 1226 AA | 5 | 5 | 100.0 |
| TOTALS | 1390 | 376 | 27.1 |

TABLE 8.2
Data Products Summary

| Scanner System or Type of Run | Number of Runs Generated |
|----------------------------------|-----------------------------|
| ERTS MSS SYCI | 565 |
| ERTS MSS SCCI | 11 |
| NASA 24 Channel | 26 |
| ERIM | 36 |
| Registrations (Overlay Runs) | 76 |
| Geometrically Corrected Runs | <u>197</u> |
| TOTAL | 911 |

8.3 Temporal Registration

The capability to digitally register multiple images of the same scene was developed by our Laboratory prior to the ERTS study. Registration of multiple ERTS passes over the same area was included in the plan of study to enable investigators to study the temporal dimension in addition to the spectral and spatial dimensions available from any one ERTS frame. Many investigators made use of this capability and in one case five passes were registered forming a 20 channel data set. Most of the registrations were three time (12 channel) combinations and were used to enable temporal classifications, change detection and relocation of objects of interest at times subsequent to the original location of the objects.

Registration of multiple images of the same scene was accomplished through use of the LARS image registration system. The registration processing operation consists of two basic operations: 1.) image correlation and 2.) registration transformation which are performed sequentially. Many factors exist which prevent exact overlay of the images. Two major errors are: (1) It is unlikely that the samples from one time were imaged from exactly the same spot as samples from a later satellite pass, thus, in general, no data exists which exactly overlays for both times even if no other errors were present; and (2) Due to changes in the scene and other "noise" sources the two images cannot be exactly correlated or matched. The registration procedure used consists of the following:

1. Initial checkpoints or matching points are manually selected in the two images to be registered using the LARS digital display. At least seven points are found and the coordinates are recorded on punched

cards. Each checkpoint consists of an ordered quadruple of coordinates:

$$P = (X_A^{(k)}, Y_A^{(k)}, X_B^{(k)}, Y_B^{(k)})$$

with X_A, Y_A being the coordinate of a point in the A or reference image and X_B, Y_B being the coordinates of the corresponding point in the B image to be registered on the A image. This step in essence removes rotational misalignment and reduces translational misregistration to 10 to 20 pixels.

2. A two dimensional least squares quadratic polynomial is generated to represent the difference in position of points in the A and B images. The polynomial is of the form:

$$\Delta X = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy$$

$$\Delta Y = b_0 + b_1x + b_2y + b_3x^2 + b_4y^2 + b_5xy$$

and the least squares solution for the coefficients is:

$$\alpha = (P^T P)^{-1} P^T \delta_x$$

$$\beta = (P^T P)^{-1} P^T \delta_y$$

Where: α, β are 6x1 coefficient vectors for ΔX & ΔY , P is the matrix $[P_{ij}]$ of powers of x and y for each checkpoint:

$$P_{ij} = X_i^k Y_i^l \text{ where } i \text{ is the number of the checkpoint,}$$

$$i=1, N; k=0,1,0,2,0,1; \quad l=0,0,1,0,2,1 \text{ for}$$

$$j=1,2,3,4,5,6 \text{ respectively. } \delta_{x,y} = N \times 1 \text{ column vector}$$

of difference between A and B coordinates,

$$\delta_{x_i} = X_{B_i} - X_{A_i}, \delta_{y_i} = Y_{B_i} - Y_{A_i}$$

This function describes an approximate registration of A and B.

3. A block image cross correlator is employed to find the remaining image displacement at the nodes of a uniform grid using the approximate registration polynomial generated in (2). The correlator implements the correlation coefficient equation:

$$R(k, \ell) = \frac{E[(A - M_A)(B_{k, \ell} - M_B)]}{\sqrt{E[(A - M_A)^2]E[(B_{k, \ell} - M_B)^2]}}$$

Where E denotes mathematical expectation, $M_{A,B}$ the mean values of A and B data blocks and the k, ℓ subscript on B denotes the shift of the B block with respect to the A block of k rows and ℓ columns. As large a set of correlations as possible is obtained within computation time constraints. The k, ℓ values at the maximum R are chosen as the correct shift to match the block from image B to the block from image A. This peak is interpolated using three point LaGrange polynomials to produce a fractional estimate of shift. The set of shifts from the correlator are added to the shift values from the original polynomial to form a new set of checkpoints.

4. A new registration polynomial is generated from the correlator produced set of checkpoints and used to actually register the images. The nearest neighbor

rule is employed to obtain points where no data exists. The A and B images are combined onto one data tape and a new data set is formed having $M+N$ channels where M is the number of channels from image A and N is the number of channels from image B.

5. The registration data tape is inspected to check image quality and registration quality. A measure of error is obtained from the residual from the least squares polynomial generation operation and this figure averages .5 of an image sample, RMS. Re-correlation of the registered images is performed to evaluate to accuracy between the points used for checkpoints.

The accuracy of registration varies with the degree of correlation between the images. Where correlation is low due to seasonal changes, RMS registration error tends to approach and exceed one pixel. For images having correlations well above .5 RMS errors in the .3 to .6 range are observed. Figure 8.5 presents the results of a test correlation of a registration of Tippecanoe County data from September 30, 1972 and June 9, 1973. The RMS Euclidian registration was .65 pixel and the RMS peak correlation was .64 and the maximum registration error 1.4 pixels. This is a typical result for agricultural area data.

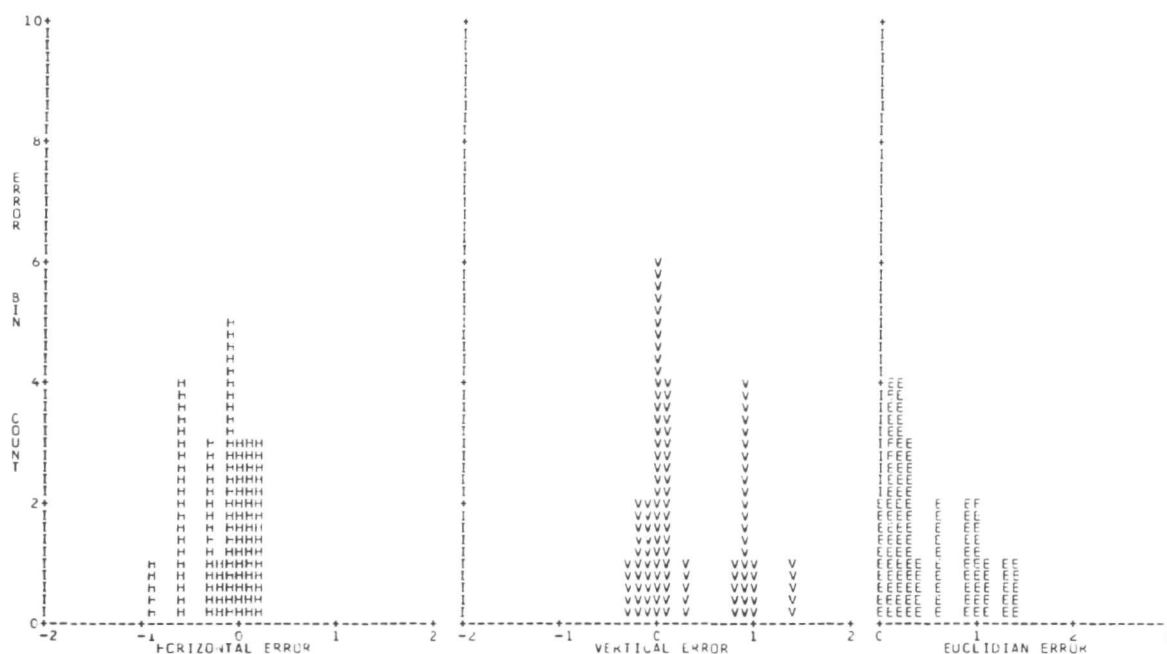
Another evaluation method consists of subtracting registered image pairs and examining the difference image for fringes and double line effects which would indicate misregistration. The difference images are also useful for change analysis of the scene. Figure 8.6 contains difference images for the September 30 - June 9 registration for the four MSS bands. No Fringing or double line effects were noted verifying the less than one pixel error estimated by the correlator. The difference images offer a test of registration accuracy whereas complete re-correlation

REGISTRATION ACCURACY EVALUATION **SEPT. 30, 1972-JUNE 9, 1973 TEMPORAL OVERLAY** **LAFAYETTE, INDIANA AREA**

***** CORRELATION RESULTS *****

| | |
|--|---------------------------------------|
| RMS EUCLIDIAN DISTANCE..... = 0.65 | AVERAGE EUCLIDIAN DISTANCE.. = 0.49 |
| RMS PEAK CORRELATION VALUE..... = 0.64 | AVERAGE PEAK CORRELATION VALUE = 0.63 |
| NUMBER OF ACCEPTABLE CORRELATIONS = 23 | NUMBER OF REJECTED CORRELATIONS = 77 |
| CORRELATION WINDOW = 32 BY 32 SQUARE | CORRELATION DELTA..... = 6 |

***** HISTOGRAMS OF CORRELATION DISTANCE ERRORS *****



RUN 72053607 CORRELATION OF CHANNEL 2 (.6-.7 μ m)

Figure 8.5 Registration Accuracy Evaluation September 30, 1972 - June 9, 1973 ERTS-1 Data Registration. Band 5 Frame each time correlated. RMS Registration Error is .65.



Difference Image for Band 4 (.5 - .6μm)



Difference Image for Band 5 (.6 - .7μm)



Difference Image for Band 6 (.7 - .8μm)



Differences Image for Band 7 (.8 - 1.1μm)

Figure 8.6 Difference images for ERTS-1 data collected over Lafayette, Indiana 9/30/72 and 6/9/73. Bright areas indicate increased reflectance over the period, gray areas relate to no-change and dark areas indicate a decrease in reflectance. Misregistration would be indicated by double lines and edges.

of the registered images requires considerable computer time.

The image registration procedure was set up on an operational basis by Summer 1973. Results using registered data are reported in several other sections of this report. The temporal registration effort is considered highly successful as it made temporal data available for machine processing over large areas and over an 18 month time span.

8.4 Data Quality Evaluation

A variety of tests were made on the system corrected CCT data throughout the year. This work was separate from the study of scene corrected data described in Section 10. The initial investigation explored the problem of horizontal striping apparent to varying degrees in the image. An example is shown in Figure 8.7. This effect is due to imperfect calibration of the six-detector array. A procedure was employed which averages all data points from each detector for each channel over a large image area, in some cases up to two-thirds of a frame. The means obtained over a large image sample should be the same and any differences are indicative of gain and offset differences between detectors. Table 8.3 contains the mean and standard deviation values obtained for ERTS frame 1016-1605000 obtained on August 8, 1972 over Southern Illinois.

Table 8.3

Data Mean and Standard Deviations for each Detector
for each Channel (MSS System Corrected Data)

| Detector | Band 4 | | Band 5 | | Band 6 | | Band 7 | |
|-----------------|--------|----------|--------|----------|--------|----------|--------|----------|
| | M | σ | M | σ | M | σ | M | σ |
| 1 | 46 | 24 | 41 | 27 | 67 | 20 | 36 | 11 |
| 2 | 48 | 25 | 41 | 27 | 68 | 22 | 36 | 11 |
| 3 | 48 | 25 | 41 | 27 | 69 | 21 | 36 | 10 |
| 4 | 47 | 25 | 40 | 26 | 68 | 21 | 35 | 10 |
| 5 | 47 | 25 | 41 | 27 | 68 | 21 | 35 | 10 |
| 6 | 47 | 25 | 40 | 26 | 66 | 21 | 35 | 10 |
| Overall Mean | 47.2 | | 40.7 | | 67.7 | | 35.5 | |

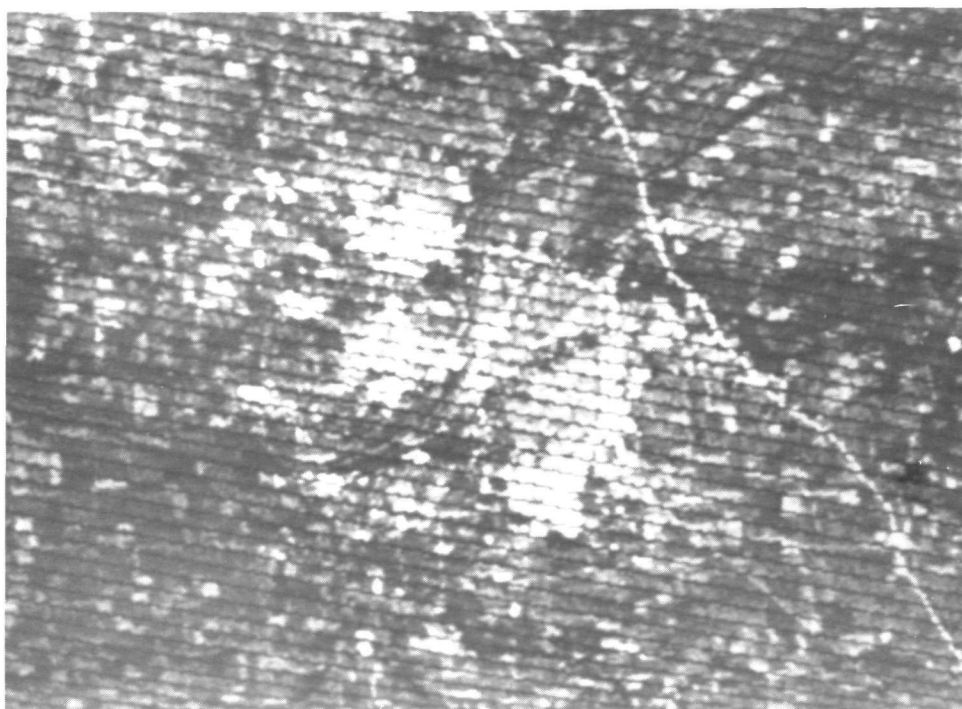


Figure 8.7 Example of striping distortion in band 4 (LARS Channel 1) observed in MSS data. Image is from Frame 1069-15585 obtained on October 19, 1972 over Lafayette, Indiana. Striping is seen primarily in band 4.

No observable striping existed in this data and the results in Table 8.3 tend to reinforce this conclusion. The maximum difference between detector means is 3 data units which occurs between detectors 3 and 6 in Band 6. The maximum difference from any 6 and overall mean is 1.7 units for detector 6 of Band 6. This is 4% of the ± 1 standard deviation range for that channel but this produced no apparent striping.

The same calculations were made on data from frame 1017-16093000 obtained on August 9, 1972 over Northern Illinois. In this data some striping could be seen in Band 4. The means and variances are presented in Table 8.4. Although essentially no differences are seen in the means, the low channel standard deviations amplify any differences that do exist. The .8 unit deviation in detector 6 of Band 4 represents 6.7 of the $\pm 1 \sigma$ data range and this level of variation evidently is observable.

Table 8.4
CCT Data Mean and Standard Deviation for
each Detector for Frame 1017-16093000

| Detector | Band 4 | | Band 5 | | Band 6 | | Band 7 | |
|--------------|--------|----------|--------|----------|--------|----------|--------|----------|
| | M | σ | M | σ | M | σ | M | σ |
| 1 | 25 | 6 | 17 | 8 | 41 | 18 | 23 | 12 |
| 2 | 25 | 6 | 17 | 7 | 37 | 20 | 22 | 13 |
| 3 | 25 | 5 | 17 | 8 | 37 | 20 | 21 | 14 |
| 4 | 25 | 6 | 16 | 7 | 37 | 21 | 21 | 14 |
| 5 | 25 | 5 | 17 | 7 | 38 | 21 | 21 | 14 |
| 6 | 24 | 6 | 17 | 7 | 39 | 20 | 21 | 13 |
| Overall Mean | 24.8 | | 16.8 | | 38.16 | | 21.5 | |

The calibration error has also manifested itself in computer classification results obtained from this and other frames under study. Horizontal strip classes are resulting in areas where there is no such structure in the actual scene. This effect is a serious distortion of the classification and is a problem which would have to be solved for any applications system.

Interchannel registration was also studied using the digital array correlator from the registration system. Band 5 was taken as the reference and correlated with Bands 4, 6, and 7 at numerous points across the frame. The correlation coefficient between channels is generally low and successful correlation can be obtained only at points where high interchannel correlation exists due to a peculiarity in the scene structure such as the edge of a body of water or a major road intersection.

Table 8.5

Mean and RMS Across Track Misregistration Estimation
of Random Points in Strips of ERTS-A LARS Run No. (72032800).

Lines 400 to 2400 by 200's
Col 100 to 1700 by 320's

| Strip | 2 | | | 3 | | | 4 | | |
|-----------------|-----------------------------|-----|------|------|-----|------|-------|-----|------|
| | # | | | # | | | # | | |
| M=Mean R=RMS | M | PTS | R | M | PTS | R | M | PTS | R |
| Band 4 | REFERENCE CHANNEL - - - - - | | | | | | | | |
| Band 5 | -.064 | 11 | .080 | .081 | 31 | .043 | -.260 | 10 | .730 |
| Band 6 | .000 | 1 | .000 | .125 | 4 | .042 | .020 | 5 | .045 |
| Band 7 | -.100 | 1 | .100 | .125 | 4 | .182 | .025 | 4 | .050 |

Table 8.5 contains averaged results of correlations judged to be reasonable estimates of across track misregistration. Correlation results were separated according to the bulk CCT strip in which they fell. The largest mean error was $-.26$ sample between Band 5 and 4 in strip 4. This would indicate that registration is outside the ± 15 meter or 18.8% tolerance stated for the system. The along track correlation results are included in Table 8.6 and here a value of $-.36$ was observed for Band 4 in strip 4. This is considerably outside the ± 3 or 3.75% meter figure quoted as are the .2 sample errors for Band 6 and 7 strip 1. The .052 figure for Band 5 strip 3 is considered acceptable. Action was not recommended on these results since the correlations are experimental and correction processing would be outside the scope of the study.

Table 8.6

Mean and RMS Along Track Misregistration Estimation
of Correlated Points in Strips of ERTS A LARS Run No. (72032800)
Lines 400 to 2400 by 200's
Col 100 to 1700 by 320's

| Strip | 2 | | | 3 | | | 4 | | |
|-----------------|-----------------------------|-----|------|------|-----|------|-------|-----|------|
| | # | | | # | | | # | | |
| M=Mean R=RMS | M | PTS | R | M | PTS | R | M | PTS | R |
| Band 4 | REFERENCE CHANNEL - - - - - | | | | | | | | |
| Band 5 | .036 | 11 | | .052 | 31 | | -.360 | 10 | |
| | | | .060 | | | .635 | | | 1.53 |
| Band 6 | .200 | 1 | | .025 | 4 | | .020 | 5 | |
| | | | .200 | | | .050 | | | .014 |
| Band 7 | .200 | 1 | | .025 | 4 | | .000 | 4 | |
| | | | .200 | | | .050 | | | .000 |

8.5 Geometric Correction

ERTS-1 Multispectral Scanner Data is received from the satellite by NASA, processed, and delivered to users recorded on computer compatible tape and in photographic form. The computer tape form of the data is calibrated and line length adjusted by NASA but no geometric corrections are applied. The system-corrected photographic products are corrected for many geometric distortions including earth rotation effects in addition to the above two corrections. Also, these images are rescaled so that the horizontal and vertical scales are the same. Thus, the digital CCT form of the MSS data contains many geometric distortions and users of this data are faced with the problem of compensating for these errors.

When digital MSS data is reproduced in image form on a standard IBM computer line printer the resulting scale factor is approximately $1'' = 22400''$ in the horizontal direction and $1'' = 25200''$ in the vertical direction. This scale differential exists in addition to the skew due to earth rotation and all other geometric errors. Similarly when this data is reproduced on a video display device the horizontal scale is 56 meters per point and the vertical scale is 79 meters per point. Correction of these geometric distortions has become highly desirable by certain researchers who require that the ERTS images exactly match maps of terrain areas under study. The techniques discussed in this note are an attempt to improve the geometric quality of the ERTS digital data for research purposes.

8.51 MSS Digital Data Geometric Characteristics

The ERTS-1 MSS system produces four spectral band digitized imagery of approximately 100 (185 km) nautical mile wide strips beneath the satellite path. The scanner has an instantaneous field of view of 79 meters and scan lines are sampled at a rate such that samples are spaced approximately 56 meters apart and successive scan lines are spaced approximately 79 meters apart as determined by the forward motion of the satellite. The image data are edited so that the along track size is approximately 96.3 nautical miles (155 km). The resulting data set consists of nominally 3240 samples horizontally (E-W) and 2340 samples vertically (N-S). The geometric distortions are due to sensor, satellite, and earth effects. The major sources of error are listed briefly here:

1. Scale Differential - This is the 56 meter horizontal versus 79 meter vertical sample ratio mentioned above. These are approximate values since sensor and satellite motion effects influence the sample rate as will be discussed below.
2. Altitude Variations - The orbit is not circular and the earth is not spherical thus the altitude varies with position in orbit about the nominal 494 N.mi. value. Changing altitude causes the 79 meter resolution to vary and the 56 meter horizontal sample spacing also varies (i.e. horizontal scale). The magnitude is of the order $\Delta X = 9.26 \times 10^4 \frac{\Delta h}{h}$ where Δh is the altitude change over the frame, h is the nominal altitude and ΔX is the change in width of the image.

3. Attitude Variations - The satellite undergoes random roll, pitch, and yaw variations due to errors in its attitude control system. Roll causes a skew in the horizontal direction of magnitude $\Delta X = h\theta_R$ where h is the nominal altitude, θ_R is the roll variation over the frame, and ΔX is the amount of horizontal skew. Pitch variations cause a change in the vertical scale by changing the vertical size of the frame. The magnitude is $\Delta Y = h\theta_p$ where θ_p is the pitch variation from the top to bottom of the frame. Yaw variation causes a variable vertical skew distortion which is difficult to simply describe.
4. Earth Rotation Skew - The Eastward rotation of the earth under the satellite path causes the area scanned for a frame to be a parallelogram skewed about 5% from square or about a 5 mile shift from top to bottom.
5. Orbit Velocity Change - The variation in satellite velocity due to the eccentricity of the orbit and non-sphericity of the earth causes a vertical scale change. The change in height of the frame due to this effect is $\Delta Y \approx 8.88 \times 10^4 \frac{\Delta V}{V}$, where ΔV is the velocity change over the frame and V is the nominal velocity.
6. Scan Time Skew - The scanning mirror takes a finite time to scan one line across the scene and in that period the satellite is moving forward. A line skew occurs which is approximately 216 meters in magnitude, i.e., one side of the scan line is 216 meters advanced along the track of the satellite than the other side.

7. Nonlinear Scan Sweep - The scanning mirror does not move evenly across the scene and the deviation from linearity is estimated to be at most 395 meters at any point across the image.
8. Scan Angle Error - The look angle from nadir causes a horizontal scale error proportional to the angle. This is a very small error since the maximum look angle is $+5.78^\circ$ and amounts to a maximum of 115 meters.
9. Frame Rotation - The orientation of the frame with respect to North is approximately 13° in the U.S.A. clockwise due to the fact that the orbit inclination at the equator is approximately 99.114° . This rotation is not considered an error; however, it is convenient to work with image products which are North-oriented.

The magnitude of most of these errors are unknown, at least by LARS CCT users at present. The major errors are the scale and skew errors. Also, rotation to North-orientation is considered highly desirable. A two step process was developed to correct this data for small areas.

8.52 Geometric Correction Algorithm

The geometric correction task was divided into two steps. Corrections that could be predicted reasonably well such as scaling and skew would be performed "open loop", i.e., without feedback from ground control points, to approximately correct the data. This approach makes improved data available to users rapidly. The second stage is a "fine" correction which uses ground control checkpoints to remove the remaining several hundred meter error in the initial correction. The coarse or initial correction would be useful to those wishing to visually

relate points on maps and ERTS data and especially to those studying the millions of rectangular North South-oriented agricultural fields which exist in certain areas. The fine correction would produce images which would exactly (within 1 pixel) match the image the checkpoints were taken from over the area that the points were taken from.

The coarse correction consists of five linear transformations which act on the entire image block. This is contrasted to a nonlinear transformation which could compensate for randomly varying scale, skew and other distortions. The ERTS image consists of discrete samples of reflected energy over a two-dimensional space. The image can be thought of as a three-dimensional array $P(i,j,k)$ where i are the rows or lines of data points, j are the columns or samples across the image and the k are the channels. The data values themselves are non-negative integers having values between 0 and 127.* The four channels are assumed to be in perfect registration in this discussion so the problem can be studied as a two-dimensional single channel image problem. The ERTS image is thus defined as an array of points P with:

$$0 \leq P(i,j) \leq 127 \quad 1 \leq i \leq 2340, \quad 1 \leq j \leq 3232$$

Transformation of this array into another array which when displayed on a certain type of output device has given geometric characteristics is the geometric correction problem.

Linear transformation of elements of a two-dimensional space into another two-dimensional space is accomplished by the linear combinations:

$$Y_1 = a_{11} x_1 + a_{12} x_2$$

$$Y_2 = a_{21} x_1 + a_{22} x_2$$

*The values in Band 7 fall in the range 0 to 63.

Or in Matrix form:

$$Y = AX$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

The physical meaning of such a transformation is depicted in Figure 8.8. The nodes of the X grid represents original ERTS samples of reflected energy from discrete points on the earth. These samples are stored as a two-dimensional array of integers. The desired samples are represented by the Y grid. These samples are oriented in a re-scaled, rotated, and deskewed coordinate system. The geometric correction process assigns radiance values to nodes in the new grid using the data available from the existing grid, i.e., the raw ERTS data. Clearly the conceptually simplest way to match a map grid to the ERTS data grid is to distort the map or its topographic coordinates to match the ERTS data. This is not practical in general because large number of maps already exist in normal topographic coordinates and users wish to match ERTS data to these maps.

The linear transformation A can correct for skew and scale errors as well as rotate the image. Note that in general no original sample exists in the new grid at the desired sample points. Thus, some form of interpolation is required to perform any geometric transformation on the data. This problem is discussed in a following section. The transformations are represented by the following matrices:

1. Scale Change

The ERTS sampling ratio is approximately 3 to 2 as determined by the ratio of horizontal samples to vertical samples for the same terrain distance. The linear transformation matrix which will change the scale of the two dimensions different amounts is:

$$M = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}$$

Note that in order to change the scale of the data some samples will have to be skipped or duplicated at some points. This can be considered elimination of information or duplication of information. Since the IFOV of the ERTS MSS is nominally 79 meters and the across track or horizontal sampling is every 56 meters redundancy already exists due to the overlap in this dimension. The vertical or along track sampling is the same as the IFOV thus no overlap exists. The question thus arises - is it preferable to eliminate partially redundant samples or duplicate independent samples to effect a scale change? It was decided that significant information would not be lost if horizontal samples were dropped. Thus the matrix to correct the horizontal scale factor from nominally 56 meters per point to 79 meters per point is:

$$M_1 = \begin{bmatrix} 1.41 & 0 \\ 0 & 1 \end{bmatrix}$$

A basic feel for what this matrix will do can be obtained by observing the coordinate limits for a total frame. At line one and column one the transformed coordinates are:

$$\begin{bmatrix} 1.41 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.41 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The resultant coordinate is rounded to the nearest integer under the nearest neighbor rule which will be discussed. At the lower right corner of the new "square" image the point 2340 lines, 2340 columns should come from the lower right corner of the original data or line 2340, column 3232. Thus when output column coordinate is 2340 the coordinate for the input is $1.41 \times 2340 = 3232$. This is not an equality since $1.41 = 79/56$ and $3232/2340 = 1.38$. The first ratio is used since it is assumed that it is more stable, i.e., the number of samples tends to be variable. This matrix is referred to as M_1 and is labeled the scanner scale correction.

2. Rotation

Rotation through an angle θ of the "squared up" image obtained from the M_1 transformation is accomplished by a standard coordinate rotation:

$$M_2 = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

The amount of rotation of the ERTS frame required to bring it square with North varies with latitude. The ERTS orbit crosses the equator with an inclination of approximately 99.119° , i.e., clockwise from North 9.119° . At the highest latitude reached (about 80°) the heading of the satellite is 90° in the Southern hemisphere, 270° in the Northern. Thus the θ varies from

9.119 to 90. The spherical trigonometric function for the required θ for rotating the data to North is:

$$\theta = 90 - \cos^{-1} \left[\frac{\sin \theta_E}{\cos \lambda} \right]$$

WHERE: θ_E is the inclination at the equator (9.119°)

λ is the latitude

The θ obtained is approximate because the heading is varying over the entire image and the orbit does not exactly have the assumed inclination. The errors involved are small; however, ground control data will be needed to remove the remaining errors.

3. Skew due to Earth Rotation

The earth is rotating inside the orbit of the satellite as the ERTS data is being scanned. The rotation results in an Eastward surface velocity which causes a skew in the resulting ERTS frame. The Eastward surface velocity beneath the satellite is approximately:

$$V_e = R_e \cos \lambda \omega_e$$

WHERE:

V_e = Velocity to East

R_e = Radius of earth $\approx 6.37816 \times 10^6$ meters

λ = Latitude of satellite

ω_e = Angular rate of the earth = $.7272 \times 10^{-4}$ radians/sec.

The satellite period is approximately 106 minutes so the angular rate is $\omega_o = 9.87 \times 10^{-4}$ radians/second. A 185.3 km (100 N.mi.) frame would be scanned in:

$$t_s = \frac{L}{R_e \omega_o}$$

$$= \frac{185300}{6.37816 \times 10^7 \times 9.87 \times 10^{-4}} = 29.4 \text{ sec.}$$

WHERE: L is the height (along track length)
of a frame

R_e is earth radius

ω_o is the orbital angular rate

The Eastward displacement of the earth during the scanning of a frame would be:

$$\Delta X_E = t_s V_e$$

For example, at 40°N latitude $V_E = 355.29$ meters/second
and the Eastward displacement would be:

$$X_E = 29.4 \times 355.29 = 10445.5 \text{ meters}$$

This is 10.445/185.3 of a frame or about 5.6% The earth rotation effect is actually acting at an angle to the scan lines due to the non-polar orbit thus the distance the bottom of the frame is actually displaced is:

$$\Delta X = \Delta X_E \cos \theta$$

If the skew correction is performed after rotation to North the cosine factor above becomes unity; however, the apparent orbital velocity is reduced by a $\cos \theta$ factor. The skew correction matrix for correction after rotation is:

$$M_3 = \begin{bmatrix} 1 & a_{sk} \\ 0 & 1 \end{bmatrix}$$

WHERE:

$$a_{sk} = \frac{LR_e \omega_e \cos \lambda}{R_e \omega_o L} = \frac{\omega_e \cos \lambda}{\omega_o \cos \theta} = \frac{.071713 \cos \lambda}{\cos \theta}$$

For a latitude of 37.5° the matrix is:

$$M_3 = \begin{bmatrix} 1 & .0586 \\ 0 & 1 \end{bmatrix}$$

4. Output Scale Factor

The three transformations given above approximately correct the ERTS image to a North-oriented image having a sampling scale of 79 meters per data point in both the horizontal (E-W) and vertical (N-W) directions. The data are reproduced in pictorial form on two different devices at the LARS laboratory. One is an IBM computer line printer and the other is a custombuilt IBM digital video display system. The line printer has a 10 column per inch print line and normally prints 8 lines to the inch down the page. This 8 to 10 aspect ratio must be compensated for if the printed image is to be "square" in scale. The matrix for this correction is:

$$M_4 = \begin{bmatrix} .8 & 0 \\ 0 & 1 \end{bmatrix}$$

The physical scale which will result if the above four transformations are applied to the data to be printed on the line printer is: 1" on the page = 25200" on the ground (denoted 1:25200). To correct this to a standard map scale of 1:24000 the scale adjustment matrix is used:

$$M_5 = \begin{bmatrix} a_s & 0 \\ 0 & a_s \end{bmatrix} = \begin{bmatrix} .9527 & 0 \\ 0 & .9527 \end{bmatrix}$$

The resulting scale can be adjusted to any value by proper choice of a_s .

The digital image display has an aspect ratio of 1:1 so the correction matrix M_4 is not needed. Also, since the scale of photographs produced from the digital display depends on the size of the print no scale adjustment is needed. The sampling scale of data prepared for the digital display is 79 meters per point in both directions and the final physical scale can be determined only after a photo print is generated. The scale of the image on the 16" (horizontal width) screen of the display is approximately 1:151000 if every screen point represents one data point.

All of the transformations can be performed at once by multiplying the matrices in the appropriate order. A 1:24000 scale line printer correction is performed by the product of the five example matrices given above:

$$M = M_1 \times M_2 \times M_3 \times M_4 \times M_5 = \begin{bmatrix} 1.03574 & .34312 \\ -.15222 & .93351 \end{bmatrix}$$

The word "approximate" was used throughout the discussion and it should be emphasized that most of the parameters used are not known accurately, thus these corrections are not exact. The sensor and satellite induced errors vary randomly over the frame thus the "rigid body" assumption implicit in the use of the linear transformation is also invalid. The accuracy of the correction is therefore unknown; however, measurements made using topographic maps indicate about a 1 to 2% scale error. This means that if a point in the data is exactly lined up with a known ground point that, in say 1000 meters, the image would be 10 to 20 meters in error from the true ground point. Figure 8.8 is a comparison of digital display images of uncorrected and



LARS digital display image of original ERTS-1 MSS system-corrected CCT data collected August 9, 1972 over Northern Illinois (Band 5, .6 - .7 μ m).

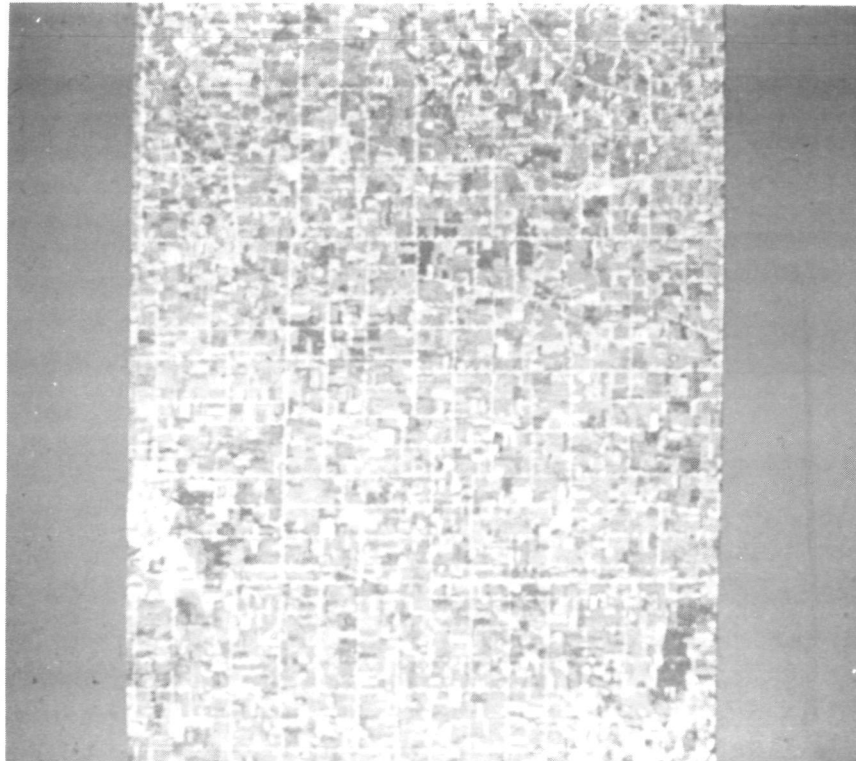


Figure 8.8 Geometrically corrected data from area shown above. Data deskewed, rotated to North orientation and rescaled such that when reproduced in image form on a computer line printer scale would be approximately 1:24000.

corrected data.

8.53 Intersample Interpolation

It can be seen from Figure 8.9 that when the geometric transformation is applied to a sampled image new samples will be needed between existing samples, i.e. where there is no data. Thus, some interpolation scheme is required to produce new samples if a uniform output grid is required. The preferred way of performing geometric transformation would be to place existing samples in the correct locations in the output image; however, this requires a randomly addressable output device with variable sample spacing. The computer line printer, LARS digital display and most other digital-to-analog image output devices have a fixed uniform point spacing so there is no way to randomly address the output image with these devices. Thus, sample interpolation is required when fixed grid output devices are to be used.

Sample interpolation can be performed in two ways: 1.) A combination of values of samples near the desired sample can be used to estimate the value at the desired point, 2.) The point nearest the desired sample location can be used to represent the value at the desired location, this is called the "Nearest Neighbor Rule". Method 1 distorts the original values of the data and it is generally assumed that the new values created this way would generate spurious multispectral vectors and would cause classification results unrelated to that of the surrounding points. Method 2 does not alter the values of the multispectral vectors so the classifications for these points will be predictable. Also, inspection of the grids in Figure 8.9 will reveal that the new point generated by the nearest neighbor rule will not be more than one sample space away from its true position in the image. The bound on the position error is:

● Original ERTS Data Grid-X
 ▲ New Transformed Grid-Y

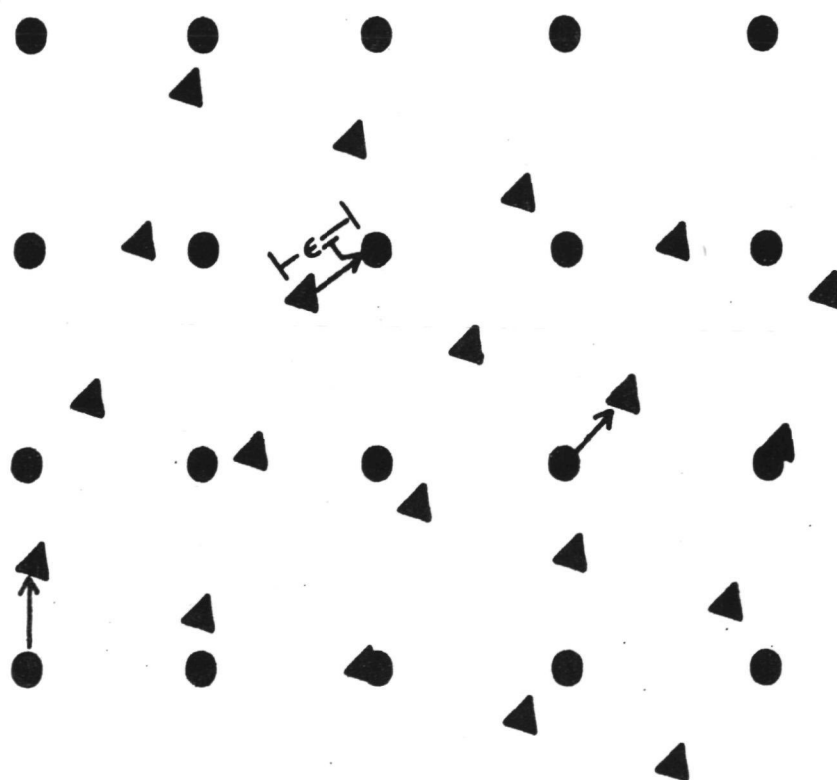


Figure 8.9 Relationship of Original and Transformed ERTS Data Points.

$$0 \leq \epsilon_T \leq \frac{1}{2} \sqrt{\Delta L^2 + \Delta C^2} = \epsilon_{MAX}$$

Where: ϵ_T = Total Euclidian Error Distance

ΔL = Line Spacing in the Data (ft or meters)

ΔC = Column Spacing in the Data (ft or meters)

For ERTS-1 data $\Delta L \approx 79$ meters and $\Delta C \approx 56$ meters thus the upper bound on the position error is 48.4 meters or 158.5 feet. The distribution of the error over the interval $(0, \epsilon_{MAX})$ would intuitively seem to be uniform for which the mean value would be $\epsilon_{MAX}/2$.

The error for each point can be computed explicitly. The locations of points required from the original data are given by the transformation:

$$X_L = f_L(y_L, y_C)$$

$$X_C = f_C(y_L, y_C)$$

Where:

$y_{L,C}$ = Line, Column Coordinates of the new Data Set.

$X_{L,C}$ = Coordinates of required points in the "old" original data set.

The new or Y coordinates are integer line and column numbers. Thus $y_{L,C} = 1, 2, \dots, N$. The $X_{L,C}$ will in general be real numbers. The error under the nearest neighbor rule will be:

$$\epsilon = X_L - [X_L]$$

$$\epsilon_L = \begin{cases} \text{If } 0 \leq |\epsilon| \leq .5 & \epsilon_L = |\epsilon| \quad \text{for lines,} \\ \text{If } .5 < |\epsilon| < 1 & \epsilon_L = |\epsilon| - .5 \end{cases}$$

$$\epsilon = X_C - [X_C]$$

$$\epsilon_C = \begin{cases} \text{If } 0 \leq |\epsilon| \leq .5 & \epsilon_C = |\epsilon| \text{ for columns,} \\ \text{If } .5 < |\epsilon| \leq 1 & \epsilon_C = 1 - |\epsilon| \end{cases}$$

where $[X]$ denotes greatest integer less than X .

For image rotation, deskewing and rescaling a linear transformation of the form:

$$X_L = a_{11}Y_L + a_{12}Y_C$$

$$X_C = a_{21}Y_L + a_{22}Y_C \text{ is used.}$$

Section 8.52 gave an example matrix for a rotation of approximately 12 degrees, rescaling to a line printer scale of $1'' = 24000''$, and deskewing 5% which is typical of operations for ERTS data. The transformation is:

$$\begin{bmatrix} X_C \\ X_L \end{bmatrix} = \begin{bmatrix} 1.03574 & .34312 \\ -.15222 & .93351 \end{bmatrix} \begin{bmatrix} Y_C \\ Y_L \end{bmatrix}$$

The distribution was evaluated using a simple program which computes the error mean and distribution for 1000 values of Y_L and 1000 values of Y_C for a total of 10^6 points. The experimental mean was .23 for each dimension which agrees well with the intuitive value of .25. The average distance error is:

$$\epsilon_T = (79 \times .23)^2 + (56 \times .23)^2 = 19.6 \text{ meters}$$

Thus, on the average about 20M or 66 feet of position error is introduced by geometric transformation of ERTS data using the nearest neighbor rule. This error is only slightly more than the 50 feet tolerance for 1:24000 scale topographic maps generated by the U.S. Geological Survey.

8.54 Geometric Correction using Ground Control Points

The correction process described above uses no ground reference points to aid in determining the values of the correction parameters. The ERTS digital aspect ratio, orbit inclination and satellite velocity are all estimated values and all are slightly in error. Improved geometric accuracy can be achieved through precise knowledge of all parameters or by finding matching points in the scene and in the data and using these points to correct the data. The second approach was investigated and preliminary results are discussed next.

An experimental precision correction was carried out in conjunction with a project funded by the U. S. Geological Survey and excellent results were obtained as determined by visual inspection. CCT data from ERTS frame 1003-18175 was first corrected for scale, rotation, and skew using techniques discussed above. The data was scaled so that when printed in pictorial form on a computer line printer the scale is approximately 1" = 24000". Easily identifiable features such as schoolyards and parks were manually located on 1:24000 topographic maps. The corresponding areas were located in the ERTS data printouts. The map used was USGS 7 1/2 minute quad: San Jose West. Thirty-six matching points were found covering a 10 x 7 1/2 mile area. The coordinate system used for the map points was the UTM system. Vertical and horizontal coordinates were measured to the nearest 10 meters and punched in standard LARS checkpoint format on cards along with the line and column coordinates for the same point in the data. These coordinates were processed by a geometric distortion function estimation program and parameters were computed to correct the remaining geometric error in the data for the given area. The data was then re-geometrically corrected to produce the final version. The results were overlayed on the topographic map to inspect the accuracy of the fit. No error could be visually observed over the 7 1/2 x 10 mile area although it is extremely

difficult to estimate locations to better than one or two pixels in ERTS-1 data. The correction function used was a quadratic polynomial with terms up to xy . A least squares fit was used to the given checkpoints. The error in estimating the checkpoints by the polynomial was .6 of a resolution element RMS.

This approach holds promise for accurately correcting ERTS type data to map coordinates. The main problem is finding matching points in the scene and the data.

8.6 Conclusions

This section described the ERTS-1 MSS data preprocessing operations which were developed and supplied to the other eight projects during the course of the Wabash Valley Study. The basic cataloging and reformatting functions were responsible for providing large quantities of MSS CCT data to LARS investigators conducting digital machine analysis of the data. This system was highly successful and indicated the desirability of a well organized reformatting phase in a digital remote sensor data analysis project.

The data quality analysis phase was operative in the early months of the study; however, as the rate of MSS CCT data received grew and the total volume of preprocessing mushroomed, detailed checking of each frame became impossible. Some examples of data quality problems from early frames are presented and discussed. Data quality analysis later in the study consisted of spot checking and analyzing user complaints. Limited corrective procedures for the striping effect were implemented late in the study but no results were obtained by the end of the study. Serious MSS data quality problems were known to exist on some frames throughout the study and the general approach can be said to be one of noting problems and avoiding use of the effected data.

Digital registration of multiple frames over the same scene was being accomplished on a routine basis by the halfway point in the study. Registration technology was developed at LARS before the ERTS study and adaptations were made to handle ERTS MSS CCT data. Temporal registration offered two capabilities to users. One was the ability to define locations of features of interest in registered data based on the locations at a reference time. For example, when several hundred agricultural fields are being studied repetatively at several times this

results in great savings in field coordinate finding. The second benefit is the availability of the temporal dimension for classification analysis and change detection. This capability was utilized extensively during the study by several study projects and results are presented in the appropriate sections.

The last task area in this project was geometric correction of CCT data. Although not defined in the original plan the need for geometric correction became very clear and steps were taken to provide a basic capability. Operational correction was offered to users by the fall of 1973 to remove the scale differential and effect of earth rotation. The frames were also rotated to North orientation. Primary output products were line printer imagery scaled to 1:24000 and CRT display imagery photos scaled by photo enlargement to various scales from 1:100000 to 1:1,000,000. The scale accuracy of the products was 1 to 2%, thus the products were not cartographic but proved to be of great value to the applications project investigators.

9.0 ATMOSPHERIC CORRECTION OF ERTS MULTISPECTRAL SCANNER DATA

9.1 Introduction

The ultimate objective of the atmospheric modelling studies conducted at LARS has been to develop algorithms which improve the capability to remotely identify earth surface features by accounting for the presence of the atmosphere. Through the processes of scattering of radiation by the particles which comprise the atmosphere and the absorption and re-emission by certain of its gaseous components the atmosphere contributes to the signal obtained by a remote sensing instrument. If one were to view as noise this addition to the data which would be obtained in the absence of an atmosphere, then our goal is to improve the signal-to-noise ratio with an appropriately designed data analysis filter.

9.2 Model Development

A physical model was adopted as the tool with which the goal could be reached. Such a model had to possess the capability to account for a surface and atmosphere both of which can absorb and scatter radiative energy and which interact between themselves subject only to the conservation of energy. The change dI_λ in an intensity I_λ as it passes through a volume characterized by an absorption coefficient $k_{\lambda a}$ and a scattering coefficient $k_{\lambda s}$ is produced by three terms: 1) the extinction (the sum of absorption and scattering) of energy from the direction of propagation of I_λ , 2) the scattering of energy into the direction of propagation of I_λ from all possible incident directions, and 3) the emission from within the volume of radiation into the direction of I_λ . Upon solving for the emergent radiance, two terms result: the

first is the contribution of the original intensity reduced by extinction within the volume, and the second is the additional radiance contributed by scattering and emission within the volume itself.

9.3 Model Evaluation

Computation of intensities is made difficult particularly by the scattering process, since the intensities to be solved for appear in both differentiated and integrated forms within the same equation, properly called an integro-differential equation. Three techniques are available to solve the radiative transfer equation in a scattering atmosphere. The iterative technique consists of repeatedly solving the transfer equation at all levels and for all directions desired until a consistent set of intensities is obtained. This would give the radiation field after one scattering process. Since multiple scatterings are possible these values then serve as input for a repeat of the iterative process for second-order scattering. The computation is repeated for successively higher orders of scattering until sufficient accuracy is obtained, as dictated by energy conservation. In the Monte Carlo technique very great numbers of individual incident beams of radiation are allowed to penetrate the scattering medium and to undergo the so-called "random-walk" to account for multiple scatterings. For sufficiently numerous repeats of this process a statistically smooth radiation field will result. In the Fourier series technique it is possible to expand the scattering phase function for a particular direction of scattering in a series whose successive terms represent higher orders of scattering. Proper combination of many such series produces the desired scattered-radiation field.

The Fourier series approach was selected and the required computational programs have been adapted to the LARS computer facility. Several physical parameters are important to the model.

- 1) Wavelength, λ : both the scattering and absorption processes are strongly wavelength dependent, a critical issue to any multispectral identification scheme.
- 2) Aerosol complex index of refraction, $m = n - ik$: the magnitude of scattered energy, given by n , and the relative amount of absorption, given by k , vary from one type of aerosol, a collection of relatively large particles of differing sizes, to another; for example, water and dust hazes have significantly different values of n and k in the visible and near-infrared portions of the spectrum.
- 3) Aerosol size distribution function, $n(r)$: under various conditions the atmosphere can contain rather more or less particles of quite small or quite large sizes; the resultant scattering of radiation is sensitive to the relative and absolute abundances of each.
- 4) Aerosol height distribution function, $n(z)$: meteorological conditions of wind and temperature determine whether particulate matter is confined near the surface or is distributed quite uniformly with height; the measured radiation at a given level in the atmosphere will be influenced by these differing conditions.
- 5) Gaseous absorption by H_2O , O_3 , O_2 , CO_2 , etc.: these several atmospheric gaseous components, some of which are quite variable with time can selectively deplete or augment the radiation in a given spectral interval, depending upon meteorological conditions.
- 6) Geometry, θ_0 , ϕ_0 , θ , ϕ : the relative location of the source of solar irradiance and of the direction of observation of emergent radiance can greatly influence measured values.

Values of radiance emergent in a given direction at a given level in the atmosphere are obtained through a series of computer

programs named SPA, SPB, SPC, and SPD. The flow of information from one to another is indicated in Figure 9.1.

SPA - computes coefficients of Legendre series for the scattering phase function of a spherical particle described by its size parameter $x = 2\pi r/\lambda$ and index of refraction $m = n - ik$.

SPB - computes coefficients of a Legendre series for the normalized scattering phase function of a unit volume (illuminated by an unpolarized, monochromatic and unidirectional beam of radiation) containing a known size distribution $n(r)$ of spherical particles all made of the same refractive index m .

SPC - computes coefficients of a Fourier series for the normalized scattering phase function of a unit volume for radiation incident at a zenith angle θ' and scattered at a zenith angle θ . The argument of the Fourier series is $\phi' - \phi$, the difference between the azimuth angles of the incident and scattered beams of radiation.

SPD - computes the intensity of the scattered radiation emerging at selected levels of a plane - parallel, nonhomogeneous atmosphere containing an arbitrary vertical distribution of ozone and water vapor concentration and/or aerosol number density, and bounded at the lower end by a Lambert ground of known reflectivity.

One of the lengthy portions of the computational procedure entails summation of the contributions of various sized particles to the total scattered energy. The computer time involved directly depends on the choice of a size interval $\Delta x = 2\pi\Delta r/\lambda$. The

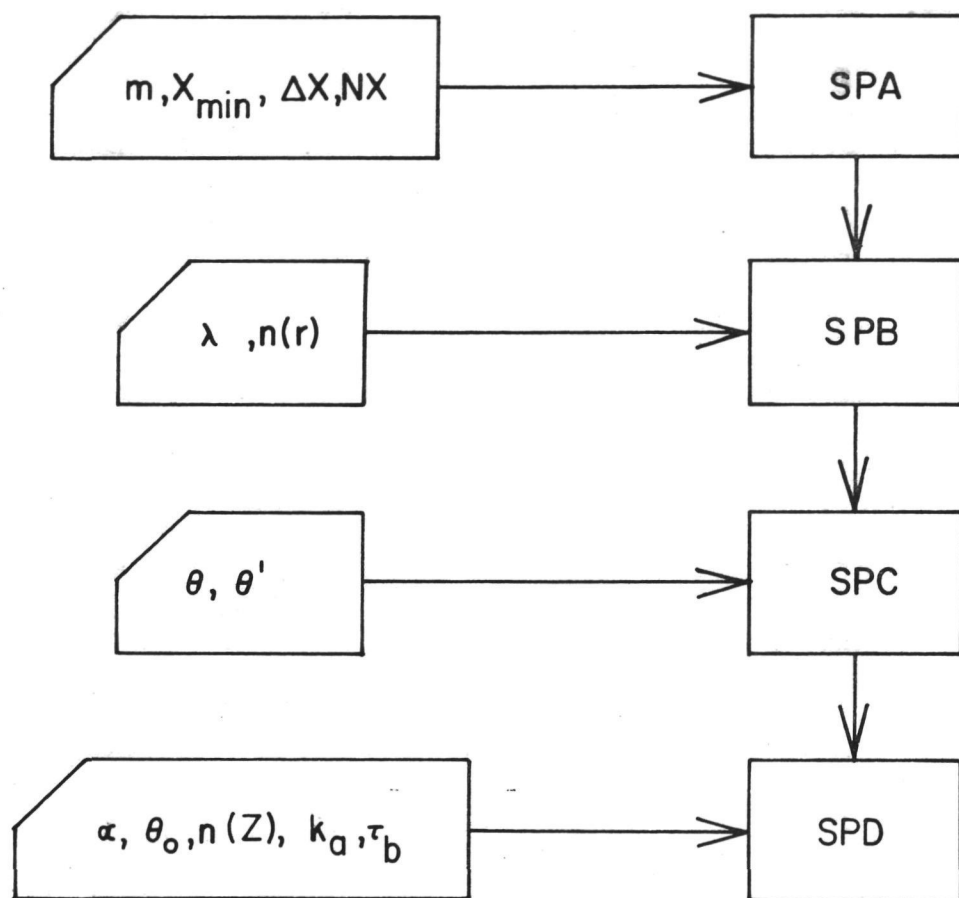


Figure 9.1 Input parameters and flow of information between the several computing programs.

accuracy of computed intensities also depends upon the resolution chosen for this numerical integration. Further, the computer time required depends upon the total size range considered, in particular upon the size of the largest particles included in the model; this is given in terms of r_{\max} , the radius of the largest aerosol particles. Again, the resultant intensities are influenced by the contributions of these large particles. Values of intensity computed for three values of surface reflectivity are shown in Figure 9.2. For this case, $\lambda = 0.55 \mu\text{m}$ and $m = 1.50 - 0.03i$, representing a slightly absorbing aerosol. As a standard for comparison the aerosol size range extends over the 0.03 to $10 \mu\text{m}$ interval and has been integrated in steps of $\Delta x = 2\pi\Delta r/\lambda = 0.2$ over these limits. The values displayed are intensities of radiation observed along nadir as a function of solar zenith angle. The importance of surface reflectivity is apparent, since the contribution of the boundary irradiance is directly proportional to α_{sfc} . As the solar zenith angle increases and the effective atmospheric path increases, intensities decrease for all α_{sfc} , approaching zero at a solar zenith angle of 90° .

In Figure 9.3, several sets of values of Δx and r_{\max} have been chosen and the computed intensities compared to the values in Figure 9.2 as percentage departures from the standard case. Many surface features have reflectivities near 0.2 , so it was chosen as a representative figure for this error study. The lowest curve for $\Delta x = 0.5$, $r_{\max} = 10\mu\text{m}$ represents a lower resolution integration over the total aerosol size range. However, the errors introduced are only a few tenths of one percent, a small price in view of the 900% reduction in computer time. If the integration increment is increased to 1.0 , the accuracy is much less as seen in the top curve. When the presence of the large particles between 2 and $10\mu\text{m}$ in radius is ignored, the remaining

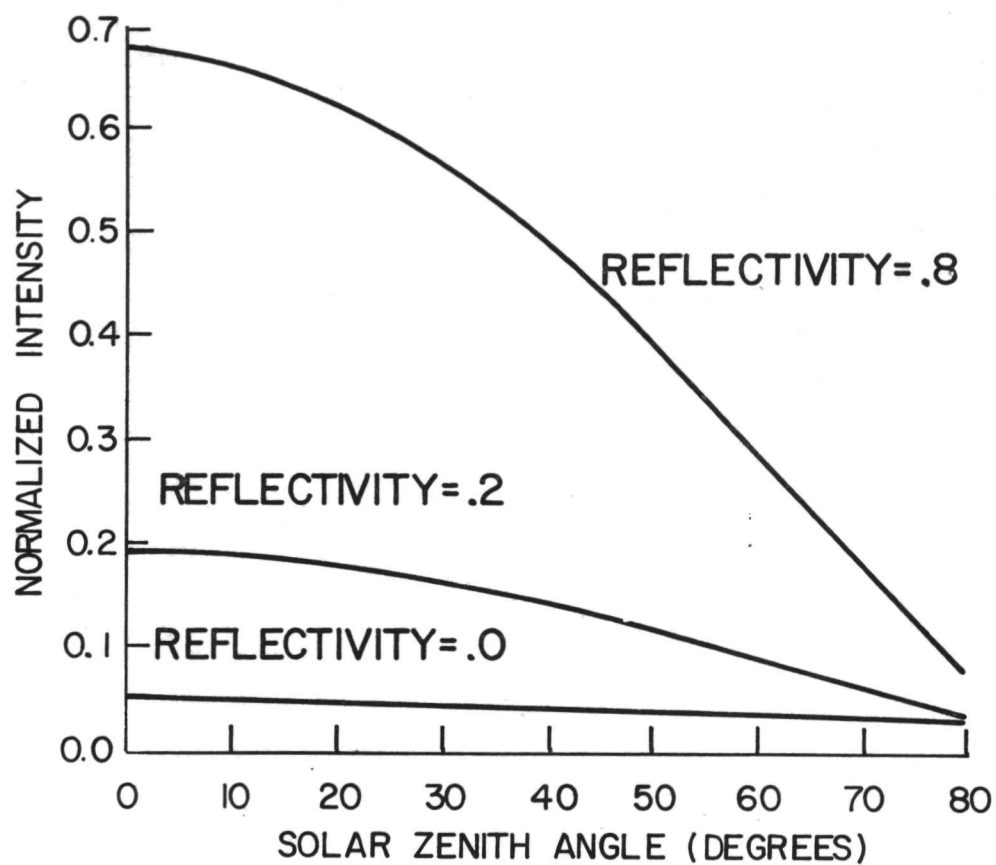


Figure 9.2 Intensities observed along nadir at the top of the atmosphere as a function of solar zenith angle for $\lambda=0.55 \text{ m}$ and haze $m = 1.50 - 0.03i$.

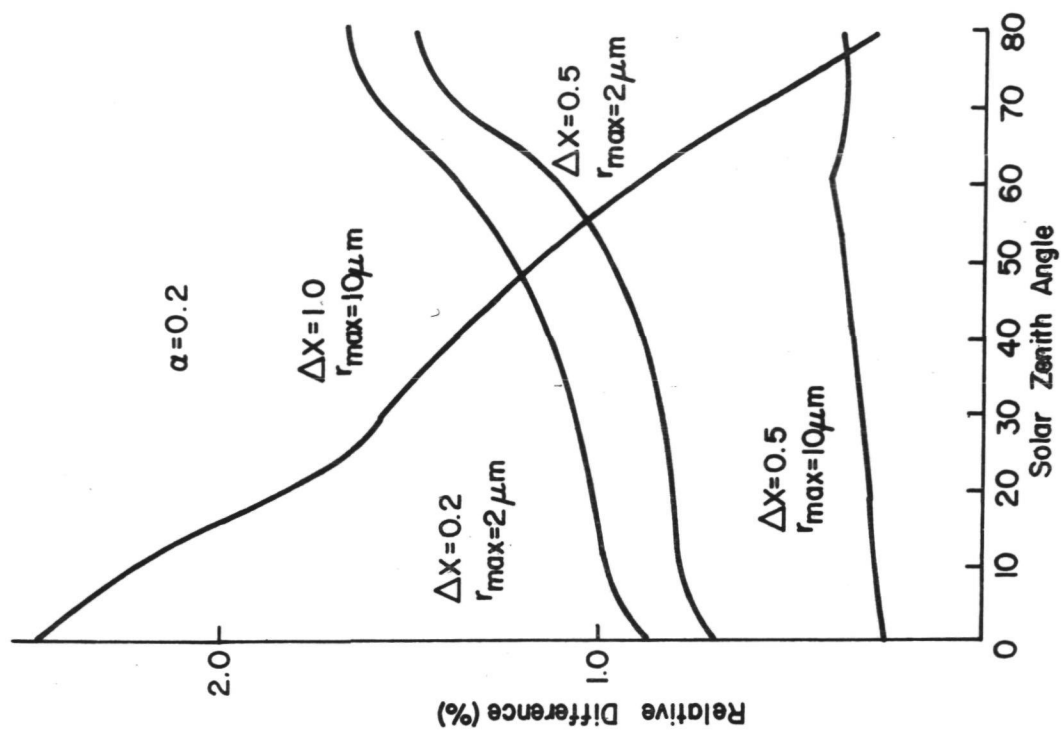
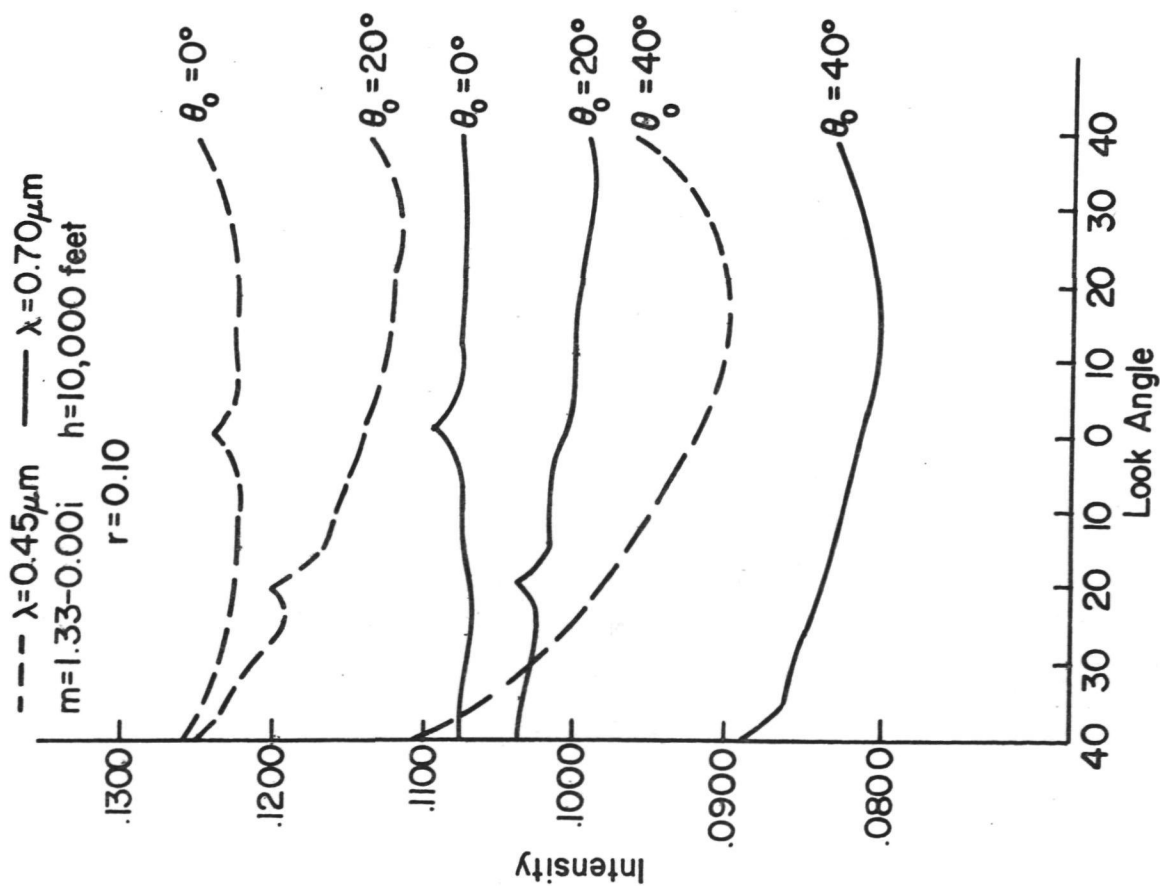


Figure 9.3 The effect of computational parameters Δx and r_{max} on reflectivity = 0.2 curve of Figure 9.2.

two curves result. Whether the higher or lower resolution Δx value is used, the errors do not exceed 1.5% regardless of the solar zenith angle. Considering that the overall limitations of the model restrict accuracy to the order to 2 to 4%, the less time-consuming case with $\Delta x = 0.5$, $r_{\max} = 2\mu\text{m}$ appears to give entirely acceptable results. These are the values of the computational parameters adopted for this study.

9.4 Parameter Specifications

Values of the other required input parameters were needed before the physical model was ready for use. The absorption coefficient for ozone, k_{O_3} , the absorption coefficient for water vapor, k_{H_2O} and the Rayleigh scattering optical thickness, τ_h , are all wavelength dependent. Values of τ_h were presented as a function of wavelength by Elterman (1968). The variation of k_{O_3} is found in the Handbook of Geophysics (1960). Detailed research on the absorptive properties of water vapor in the portion of the spectrum in which the ERTS MSS is receptive, however, has only recently been conducted.

MacDonald (1960) used the broad-band laboratory results of Fowle to determine absorptivities for water vapor in the visible and near-infrared. The four applicable absorption bands identified by Fowle are listed in Table 9.1. MacDonald considered in absolute units the solar spectrum of the energy that enters the atmosphere and determined the amount of flux in each of the four water vapor absorption bands. Using Fowle's absorption data, he was then able to conclude how much flux was absorbed in each band. The ratio of the amount absorbed to the total amount in each band is the absorptivity, which is related to the absorption coefficient. The absorptivities and coefficients for each water vapor band are also given in Table 9.2.

Table 9.1, Absorption Parameters in the Visible and Near-Infrared Spectrum

| Band Designation | Wavelength Limits | Absorptivity | Absorption Coefficient |
|------------------|-------------------------|--------------|------------------------|
| α | 0.70-0.74 μm | 0.02 | 0.04 |
| 0.8 μ | 0.79-0.84 | 0.025 | 0.05 |
| ρ | 0.86-0.89 | 0.09 | 0.185 |
| ϕ | 1.03-1.23 | 0.10 | 0.211 |

Because of the way in which MacDonald computed the absorbed flux, the coefficients have to be considered as average values over the entire bandwidth of each band.

Selby and McClatchey (1972) recently published a high-resolution study of water vapor absorption. In their atmospheric transmittance model, an absorption index for water vapor is presented every 5 cm^{-1} . By a chain of computations, each index is related to an absorption coefficient. Because of the higher resolution and more recent vintage, the results of Selby and McClatchey were adopted for use in the physical model.

Since the model requires a single value of wavelength as input, a spectrally averaged absorption coefficient for ozone, \bar{k}_{O_3} , and water vapor, $\bar{k}_{\text{H}_2\text{O}}$, and a spectrally averaged Rayleigh scattering optical thickness, $\bar{\tau}_b$, must be derived. In direct correspondence with the averaging effect of the sensor's optical system on the signal produced in each band, \bar{k}_{O_3} , $\bar{k}_{\text{H}_2\text{O}}$ and $\bar{\tau}_b$ were found using

$$\bar{k}_{\text{O}_3}(i) = \frac{\int_0^\infty k_{\text{O}_3}(\lambda) E_i(\lambda) d\lambda}{\int_0^\infty E_i(\lambda) d\lambda} \quad i = 1, 2, 3, 4$$

$$\bar{k}_{H_2O}(i) = \frac{\int_0^{\infty} k_{H_2O}(\lambda) E_i(\lambda) d\lambda}{\int_0^{\infty} E_i(\lambda) d\lambda} \quad i = 1, 2, 3, 4$$

$$\bar{\tau}_b(i) = \frac{\int_0^{\infty} \tau_b(\lambda) E_i(\lambda) d\lambda}{\int_0^{\infty} E_i(\lambda) d\lambda} \quad i = 1, 2, 3, 4$$

where $E_i(\lambda)$ is the optical filter response of the i^{th} ERTS band. These computations were made using the average ERTS MSS filter response curves shown in Figure 9.4. The results are tabulated in Table 9.2.

Table 9.2. Absorption Coefficients and Rayleigh Scattering Optical Thickness For ERTS MSS Channels

| ERTS Band | \bar{k}_{O_3} | \bar{k}_{H_2O} | $\bar{\tau}_b$ |
|-----------|-----------------|------------------|----------------|
| 4 | 0.0826 | 0.0 | 0.1028 |
| 5 | 0.0689 | 0.0 | 0.0514 |
| 6 | 0.0 | 0.0184 | 0.0298 |
| 7 | 0.0 | 0.1255 | 0.0147 |

To complete the complement of required input parameter values, a size distribution function for the atmospheric particulate matter was necessary. A discontinuous power law function as advocated by Bullrich (1964) was decided upon.

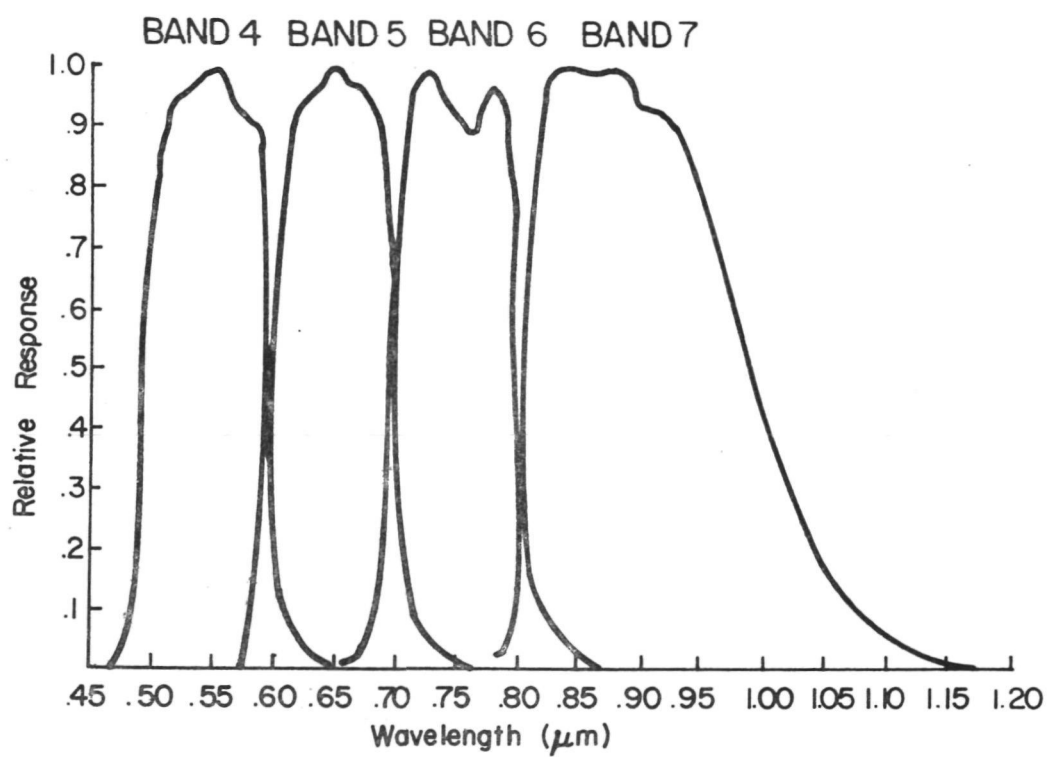


Figure 9.4 Filter response curves for the four MSS bands of ERTS.

Its form is

$$n(r) = C \quad r_{\min} \leq r \leq r_m$$

$$n(r) = C \left(\frac{r_m}{r} \right)^4 \quad r_m \leq r \leq r_{\max}$$

where $n(r)$ is the number of particles per unit volume per one micron radius interval at radius r , r_{\min} is the radius of the smallest particle included in the model, r_{\max} is the largest permissible radius, and r_m is some intermediate value.

By fixing the reflectivity of the surface in the model at a constant value, the effect of a change in the value of an atmospheric parameter could be studied. Because of its high variability and expected high degree of influence on the total transmittance of the atmosphere, water vapor content was chosen as the parameter. Figure 9,5 illustrates the variation of the upward-travelling radiance with the solar zenith angle for each of the four ERTS MSS bands at the top of the atmosphere. These results of the atmosphere model are based on the vertical distribution and content of ozone, aerosol, and water vapor for an average mid-latitude Summer atmosphere. This information is available in McClatchey et al. (1972). The driving force for the model is the solar flux density, and the fixed surface reflectivity is 20%. The solar flux received by an ERTS MSS band was found by using the same averaging scheme that was used for the absorption coefficients and the Rayleigh scattering optical thickness. The wavelength dependent solar irradiance incident at the top of the atmosphere as found in the Handbook of Geophysics (1960) was averaged to yield the driving force for each MSS band. Since the solar flux is highest in the spectral interval of the first band, the upward radiance received

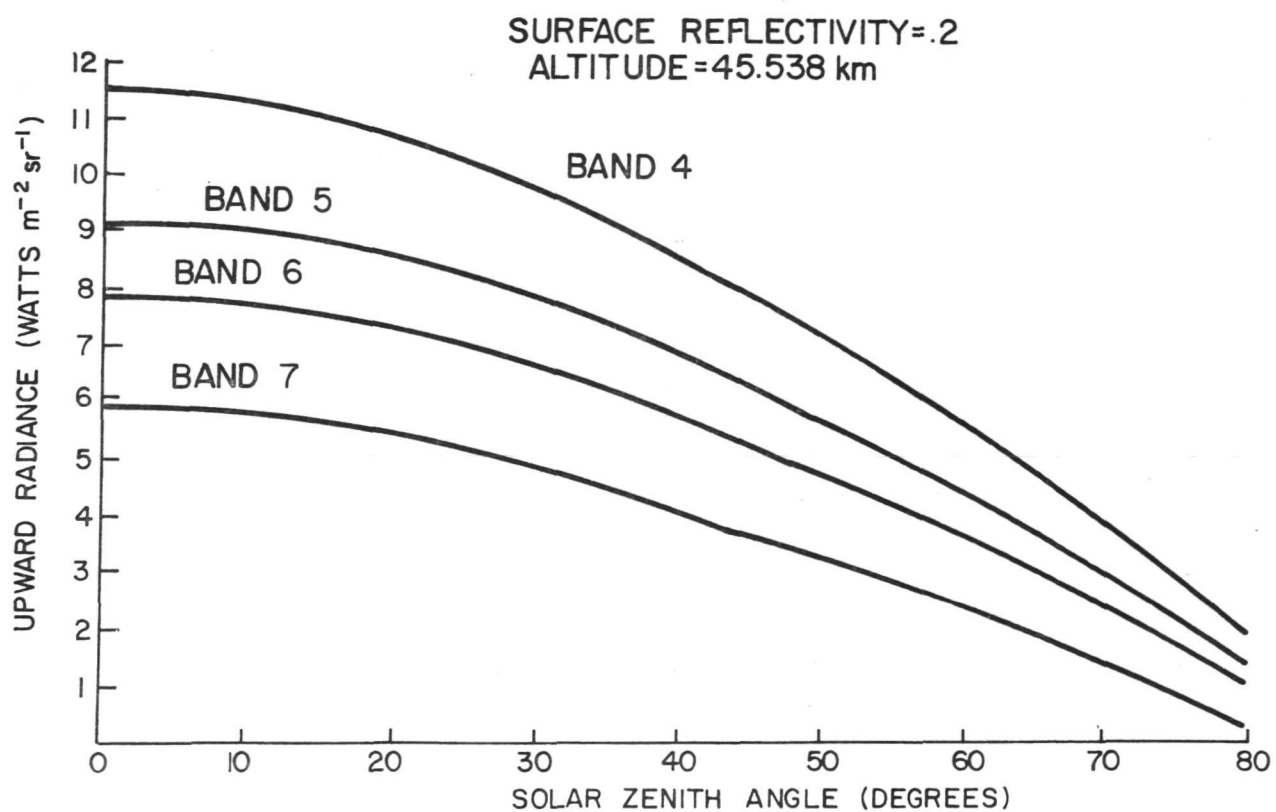


Figure 9.5 Upward radiance at the top of the atmosphere for each ERTS MSS band versus solar zenith angle. Average mid-latitude summer conditions apply.

by that band is higher than either of the other bands. Also, as the solar zenith angle increases, the upward radiance diminishes as expected because of the extra atmosphere through which the solar flux must pass.

The absorption spectrum of water vapor in the visible and near-infrared consists of weak bands located in the parts of the spectrum in which only MSS bands 6 and 7 are receptive. Thus, the radiances recorded by bands 4 and 5 are unchanged if the atmospheric water vapor content is altered and if all other things remain constant. Two additional cases of the atmospheric model have been run. In one, the water vapor content is decreased to one-half of the average mid-latitude summer atmosphere amount, and in the other, the content is reduced to zero.

Figure 9.6 displays the upward radiance reaching the MSS Band 6 sensor for the three cases of no water vapor, 1/2 the average amount, and the average mid-latitude summer atmosphere amount. At a solar zenith angle of 0° , there is only an 8.5% change in the value of the emerging radiance. Figure 9.7 is a similar graph except for the MSS Band 7 sensor. For the case of the sun directly overhead, this time there is an 80.5% change in radiance.

In conclusion, it was determined that the ERTS MSS Band 7 data is to a high degree influenced by the water vapor content of the atmosphere. Consideration of the meteorological conditions at the time of an ERTS overpass in this light appears to be essential for maximum classification accuracy to be achieved.

9.5 Application Example

Lee, Ogle, and DeKalb Counties in Northern Illinois were chosen as a case study to investigate the atmosphere's effect on

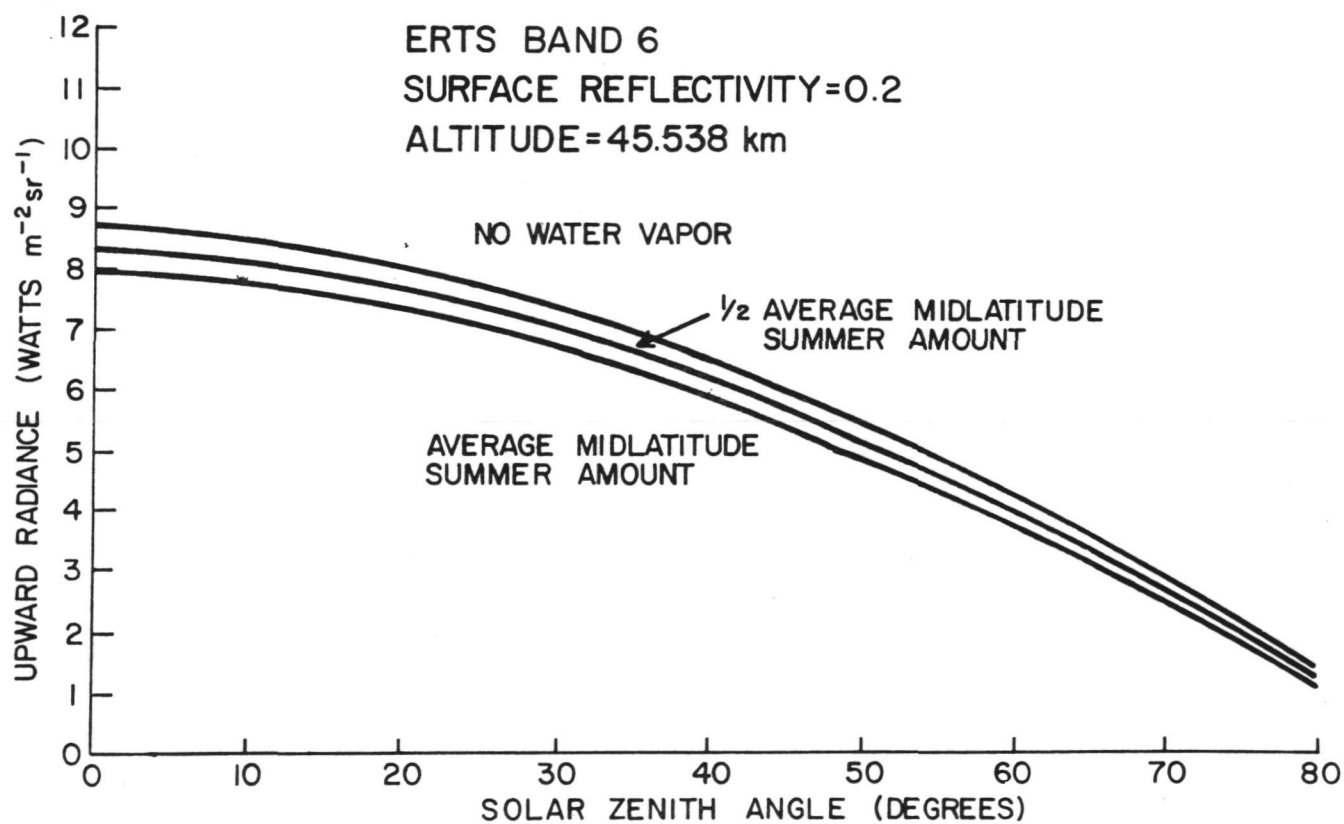


Figure 9.6 Upward radiance at the top of the atmosphere for ERTS Band 6 versus solar zenith angle for several values of total atmospheric water content.

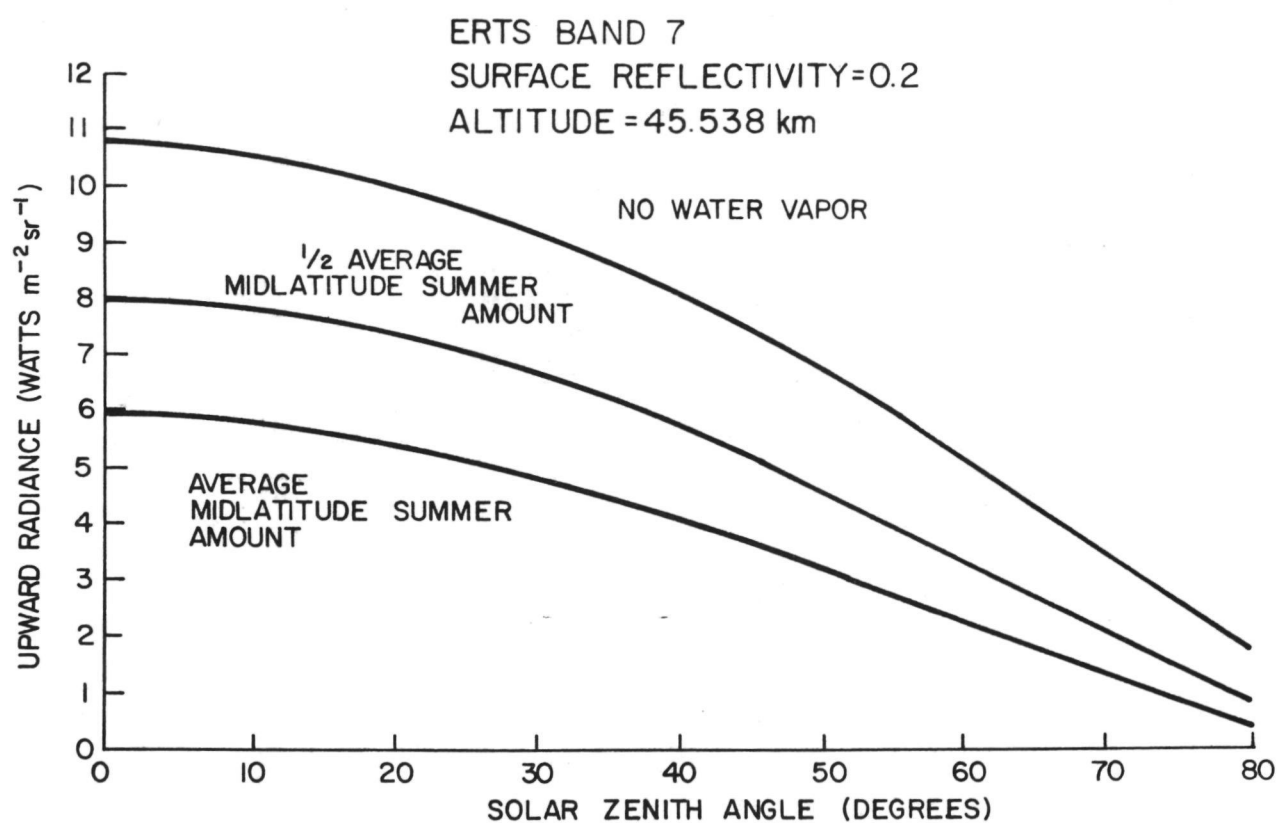


Figure 9.7 Same as Figure 9.6 except for ERTS Band 7.

classification accuracy. Meteorological data were gathered from the National Meteorological Center for the area at the time of the overpass on August 9, 1972. An upper-air radiosonde sounding from Peoria and ground observations from Peoria, Rockford, and O'Hare Airport were all received. From them, vertical distributions of ozone, water vapor, and aerosol concentration were deduced. Based on climatological data for Northern Illinois, a complex index of refraction of $1.33-0.0i$ corresponding to a water-base aerosol was chosen. Land-use information and spectrophotometer data for the prominent surface crops in the three county area were found. By weighting the crop reflectance in a band with the percentage of land occupied by that crop in August, 1972, a spectral surface reflectivity was computed. With the average absorption coefficients, the average Rayleigh scattering optical thickness, the surface reflectivities, and the vertical distributions, the atmospheric model was run for each of the four ERTS bands. The results are tabulated in Table 9.3.

Table 9.3. Northern Illinois Reflectivities and Atmospheric Transmissivities

| ERTS BAND | SURFACE REFLECTIVITY | ATMOSPHERIC TRANSMISSION (%) |
|-----------|-------------------------|---------------------------------|
| 4 | 0.1435 | 105.69 |
| 5 | 0.1096 | 89.38 |
| 6 | 0.4135 | 75.52 |
| 7 | 0.4718 | 61.90 |

In Band 4, 105.69% of the radiant energy leaving the surface reached the sensor. With this data in hand, work is proceeding to determine what the impact actually is on the accuracy with which surface features can be identified.

List of References

Bullrich, K., "Scattered Radiation in the Atmosphere and the Natural Aerosol", in Advances in Geophysics, H. E. Landsberg and J. Van Meighan, Editor, Vol. 10, Academic Press, New York, pp. 99-260, 1964.

Elterman, L., UV, Visible, and IR Attenuation for Altitudes to 50 km, 1968, AFCRL, Environmental Research Papers, No. 285, AFCRL-68-0153, 1968.

Handbook of Geophysics, The MacMillan Company, New York, 1960.

MacDonald, J. E., "Direct Absorption of Solar Radiation by Atmospheric Water Vapor", Journal of Meteorology, Vol. 17, No. 3, pp. 319-328, 1960.

McClatchey, R. A., Fenn, R. W., Selby, J.E.A., Volz, F. E., and Garing, J. S., Optical Properties of the Atmosphere (Third Edition), AFCRL, Environmental Research Papers, No. 411, AFCRL-72-0497, 1972.

10.0 Comparison of System Corrected and Scene Corrected CCT Data

10.1 Introduction

This study was initiated to investigate the quality of the scene corrected data products from the ERTS-A satellite. Since the scene corrected data has been geometrically corrected, it was thought that this data could be utilized in a data overlay scheme to geometrically correct the system corrected digital data products. This procedure would allow a researcher to maintain the radiometric quality of the system product while gaining the geometric accuracy of the scene corrected data products. The geometric accuracy would be of considerable use in improving acreage measurements of different crop types by utilizing the overlayed (system corrected upon scene corrected) data set for the classification.

The original plan of attack was to use data from the Wabash Valley Area. Scene corrected data from these areas was ordered in October 1972 and February 1973. Neither of these data products were received in time for this investigation. However, one frame of scene corrected data was received (via USDA) from the Missouri Area South of St. Louis (Scene I.D. E-1071-16111-601). This scene was thus used to complete this study comparing the system and scene corrected data sets.

Another aspect of this project was the investigation of the processing equipment hardware and software used in generating the scene corrected data products. None of the references located gave any information that was not already found in the ERTS Data Users Handbook. This document does not present much information beyond the block diagram of the system.

10.2 Data Product Description

Table 10.1 lists the basic characteristics of the data products after they were reformatted to LARS data storage tapes

Table 10.1 Details of Data Used in Study

| | <u>RUN</u> | <u>LINES</u> | <u>COLUMNS</u> | <u>DATA CHANNELS</u> | <u>INFORMATION BITS</u> |
|------------------|------------|--------------|----------------|----------------------|--------------------------------------|
| System Corrected | 72063200 | 2340 | 3232 | 4 | Bands 4,5,6:7 bits Bands 7:6 bits |
| Scene Corrected | 72063201 | 4096 | 2204 | 4 | 7 bits all channels |
| | 72063202 | 4096 | 2204 | 4 | 7 bits all channels |

Since the scene corrected data is digitized to many more data points, it was reformatted onto two tapes (the left and right halves of the frame). The only other data products available for this frame were the system corrected photographic images.

10.3 Description of Scene Corrected Data

As seen in Table 10.1, the scene corrected frame was reformatted into two LARS data storage tapes: Run 72063201 contains the left half of the scene and Run 72063202 contains the right half. As seen in Figure 10.1, which is a photograph taken from the LARS digital display, this data contains the boundary information of the imagery including the grayscale. The grayscale portion of this data product was not utilized in this study; however, it could be of some use in future studies.

When histogramming portions of the scene controlled data, care must be taken not to include the boundary areas of the frame. Histograms were calculated for an area of the right hand half of this frame. Results from this output indicate that there are 127 levels in 7 information bits for all four channels. The ERTS Data Users Handbook gives this specification for the scene corrected data computer compatible tapes.

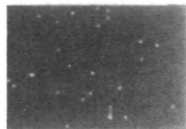
10.4 The Quality Study

Two main areas of quality were studied in this investigation. First, the geometric quality was investigated by looking for any obvious discrepancies in the scene corrected data on the digital display. Secondly, the radiometric quality was also investigated by using the LARS digital display and the LARSYS statistics program to compare the scene corrected data to the system corrected data. The results of these investigations are contained in the form of computer output and photographs.

In studying the geometric quality of the scene corrected data, attention was directed towards any obvious geometric anomalies or discontinuities shown within the digital display field outlines. In Figures 10.2, 3, 4, 5 and 6 are seen horizontal or lateral shifts in the data. Also, in Figures 10.4, 5 and 6 in the right-most box are contained what appears to be a vertical shift of the data. The exact lines or columns of the discontinuities were not recorded for Figures 10.4, 5 and 6. For Figures 10.2 and 3, the lines were recorded for these discontinuities. For Figures 10.2, the shift occurred at line 1536 and for Figure 10.3 at line 1024. These lines correspond to the boundaries of the scene corrected processing subsets. Each frame is processed in 64 different subsections and since there are 4096 lines in the frame, there are 512 lines in each subsection (i.e., $4096/8$). 1536 and 1024 are multiples of 512 and thus correspond to boundaries between two processing subsections. This fact could also explain the apparent non-uniform shifts of the data along one of these lines of discontinuity.

The radiometric quality was investigated by two means. First obvious discrepancies were located visually using the digital display on both the system corrected and the scene corrected data. Secondly, the system corrected data was compared to the scene corrected data by choosing "training" fields in the same areas on

W089-001
TS E-1071-16104-7 01



W089-001



Figure 10.1 Image of border area data of scene corrected CCT data of Frame E-1071-16111-601.

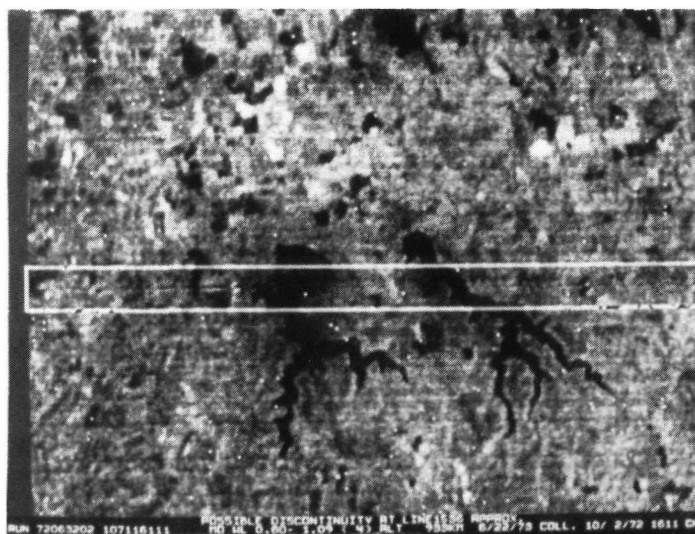


Figure 10.2 Scene corrected image showing lateral discontinuity. White rectangle was added to enclose the discontinuity. Same Frame as above. Band 7.

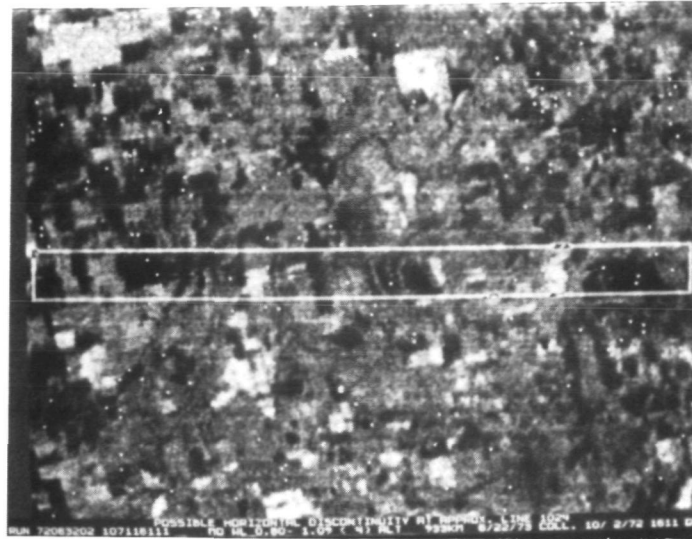


Figure 10.3 Horizontal discontinuity example in scene corrected data Frame 1071-16111-601. Band 7.

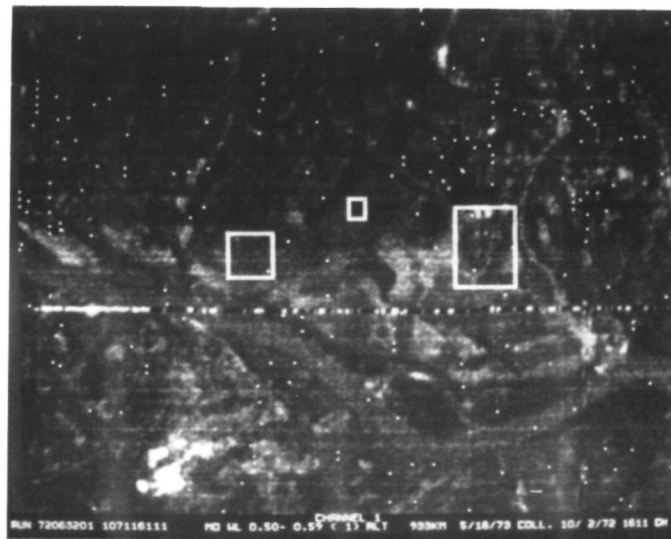


Figure 10.4 Scene corrected data showing both horizontal and vertical discontinuities. Same Frame as above. Note saturation points (white dots) throughout these images also. Band 4.

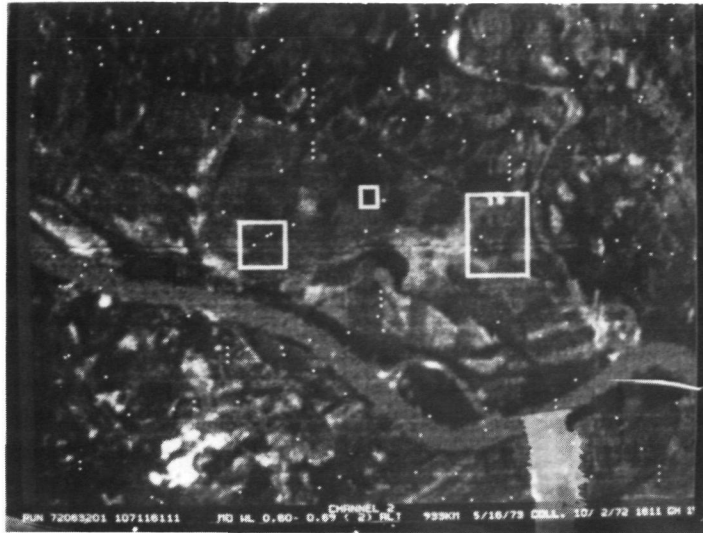


Figure 10.5 Scene corrected data showing horizontal and vertical discontinuities. Frame 1071-16111-601. Band 5.



Figure 10.6 Horizontal and vertical discontinuity examples in scene corrected data. Same Frame as above. Band 6.

both data sets. By performing computer statistical analysis on these fields in each data type, a comparison can be made. In choosing the "training" fields, water was used mainly because of its large size and ease of identification. Also two agricultural fields were chosen that could be identified in both data sets.

Referring to Figures 10.1 through 10.6 and 10.10 through 10.12, it is obvious that there are many points in the scene corrected data which are saturated (i.e., the white spots). This fact is also indicated in the histograms for the left hand portion of the scene controlled data. These saturations are indicated by the large peak at level 127 on the histograms for run #72063201. Similar peaks did not occur in the right hand half of the frame histograms. Also according to the data range only Band 7 has points at the maximum count of 128. However, from the photographs in Figures 10.2, 3 and 10 through 12, it is evident that saturated points also exist in this portion of the frame also. The only explanation available for the histogram results not showing these "bad" points, is that the histogram interval of every third point somehow missed these points.

The scene processed data is pictured in Figures 10.7 through 10.9. Note that this data appears consistent and free from saturated points. Also histograms of this run do not indicate any irregularities.

Several "training" fields were used in the statistics analysis. These fields were chosen, using the digital display, for runs 72063200 and 72063202. Two reservoirs with 5 "fields" were grouped into the class WATER1 and the three river "fields" were grouped into class WATER2. The two fields were treated as separate classes.

The results for the reservoir areas is compared by checking the correspondence of the statistics. The means and standard



Figure 10.7 System corrected CCT data from same area as Figures 10.4 thru 10.6. Note absence of discontinuities and saturation. Band 4.



Figure 10.8 Same as above for band 5.



Figure 10.9 Same as Figure 10.7 except for band 6.



Figure 10.10 Scene corrected data showing test areas from water bodies. Band 7.

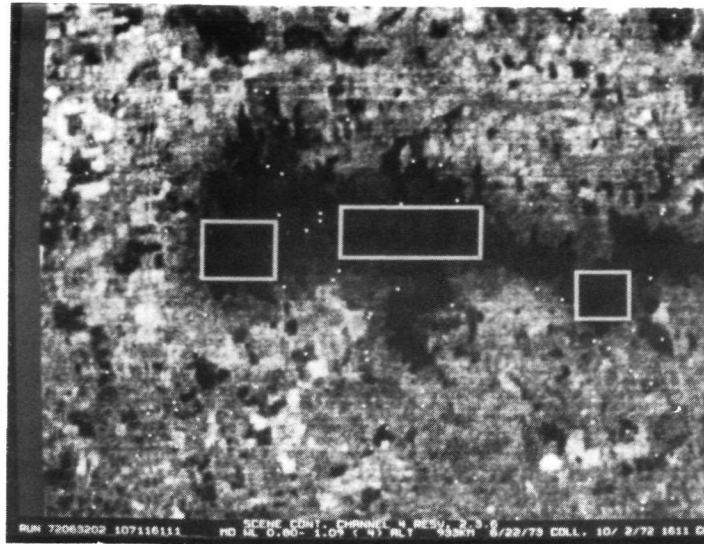


Figure 10.11 Test blocks from scene corrected data.

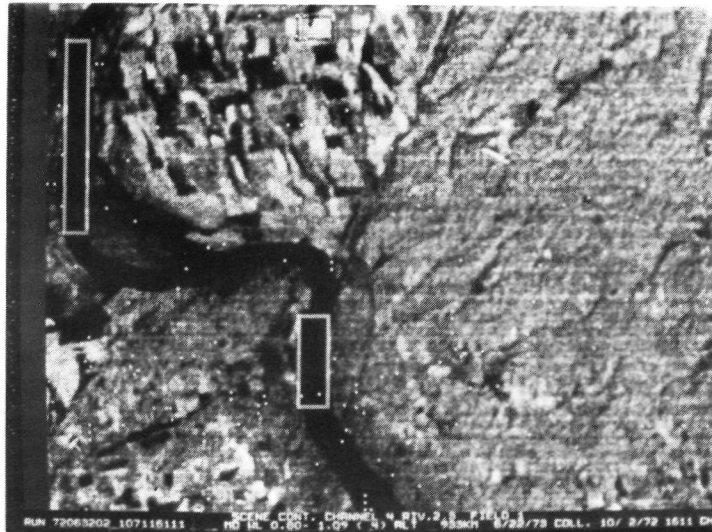


Figure 10.12 Test blocks in scene corrected data. Mississippi River is seen in the left half.

deviations for the test blocks are presented in Table 10.2. For field RESV.1 the means of Bands 4 and 5 are within a standard deviation of each other. However, the means for Bands 6 and 7 are not within three standard deviation, of a common value.

Band 7 especially contains a large discrepancy. Part of this discrepancy is because Band 7 contains only 63 levels or 6 bits for the full range in the scene processed data while the same channel in the scene processed contains 127 levels for the same irradiance range. These same general discrepancies are contained in all of the fields in the WATER1 class. Here again channels 3 and 4 are off from one another by a large amount.

In the WATER2 class, which is made up of fields from the Mississippi River, the first two channels correspond within their standard deviations while Band 7 is still a factor of 4 different. Band 7 for the system corrected data has a mean of 3.43 while Band 7 for the scene corrected data has a mean of 16.63 for the class WATER2.

For the agricultural field FLD1, Bands 4, 5, and 6 are within their respective standard deviations. Band 7 in this field has a mean of 30.00 in the system corrected data with a standard deviation of 2.26 and a mean of 50.10 with a standard deviation of 5.24 in the scene corrected data. If the mean and standard deviation of the scene processed data were multiplied by 2 and thus normalized to 127 different levels, then the Band 7 data would be within their standard deviation for the two data sets. Similar results would be obtained with FLD2 class but on these, the Band 7 results would not be within a standard deviation of one another for the two data types.

TABLE 10.2 Means and standard deviations from test blocks in scene corrected and system corrected MSS data (Frame E-1071-16111-601).

| CLASS | FIELD | SCENE CORRECTED BAND | | | | SYSTEM CORRECTED BAND | | | |
|---------|------------|----------------------|-------|-------|-------|-----------------------|-------|-------|-------|
| | | 4 | 5 | 6 | 7 | 4 | 5 | 6 | 7 |
| WATER 1 | RESV 1 | \bar{m} 24.24 | 21.30 | 18.24 | 17.37 | 23.56 | 16.59 | 8.49 | 1.44 |
| | | σ 2.13 | 3.05 | 3.29 | 2.58 | 2.23 | 2.26 | 2.50 | 1.00 |
| | RESV 2 | 20.96 | 16.73 | 16.01 | 16.05 | 20.82 | 12.62 | 7.60 | 1.24 |
| | | 2.75 | 3.21 | 1.04 | 1.03 | 0.97 | 0.85 | 1.02 | 0.60 |
| | RESV 3 | 21.40 | 16.78 | 16.30 | 16.02 | 21.43 | 13.02 | 8.44 | 1.32 |
| | | 3.18 | 2.25 | 2.95 | 0.81 | 0.86 | 0.67 | 0.95 | 0.59 |
| | RESV 4 | - | - | - | - | 17.43 | 9.78 | 8.43 | 2.82 |
| | | - | - | - | - | 1.99 | 2.37 | 6.12 | 3.90 |
| | RESV 5 | 22.14 | 18.51 | 25.19 | 19.99 | 22.53 | 15.22 | 7.41 | 1.18 |
| | | 2.79 | 2.73 | 8.55 | 4.32 | 1.35 | 0.86 | 1.13 | 0.54 |
| | RESV 6 | 21.78 | 17.89 | 17.73 | 16.64 | 21.69 | 13.84 | 12.68 | 2.05 |
| | | 1.11 | 3.71 | 5.80 | 0.70 | 1.15 | 0.95 | 9.22 | 1.42 |
| | ALL FIELDS | 22.87 | 19.31 | 18.81 | 16.70 | 22.50 | 15.10 | 8.46 | 1.47 |
| | | 2.84 | 3.56 | 5.53 | 2.67 | 2.38 | 2.67 | 3.28 | 1.28 |
| WATER 2 | RIVER 1 | 23.15 | 20.79 | 18.73 | 16.55 | 25.48 | 21.64 | 15.57 | 3.61 |
| | | 1.21 | 1.05 | 1.15 | 1.04 | 1.53 | 2.11 | 2.25 | 1.78 |
| | RIVER 2 | 24.76 | 22.00 | 18.99 | 16.58 | 26.03 | 22.19 | 15.91 | 3.30 |
| | | 5.17 | 3.25 | 2.41 | 1.46 | 0.96 | 1.09 | 1.24 | 0.83 |
| | RIVER 3 | 23.59 | 21.07 | 19.11 | 16.82 | 25.96 | 22.02 | 15.85 | 3.27 |
| | | 3.89 | 3.91 | 4.05 | 3.99 | 0.94 | 0.99 | 1.11 | 0.64 |
| | ALL FIELDS | 23.91 | 21.35 | 18.93 | 16.63 | 25.77 | 21.91 | 15.75 | 3.43 |
| | | 3.94 | 2.95 | 2.62 | 2.27 | 1.27 | 1.63 | 1.75 | 1.33 |
| FLD 1 | FIELD 1 | 22.50 | 17.28 | 42.14 | 50.10 | 23.36 | 14.47 | 49.18 | 30.00 |
| | | 1.15 | .83 | 3.56 | 5.24 | 1.27 | 1.02 | 3.50 | 2.26 |
| FLD 2 | FIELD 2 | 21.39 | 17.70 | 40.22 | 45.83 | 23.65 | 14.96 | 47.33 | 29.46 |
| | | 0.89 | 5.50 | 3.55 | 4.61 | 1.19 | 1.17 | 4.11 | 3.61 |

(Standard deviation appears below the mean for each band and channel)

10.5 Conclusions

From studying the histogram plots of the water classes, it seems that the system corrected data contains a one-sided Gaussian shape in Band 7 due to very low energy level. It seems, however, that the scene corrected data is shifted nonlinearly in the lower energy ranges towards the high energy values. This could be caused by operating the photographic portion of the system in a nonlinear region of the film.

It appears that in general the scene corrected data (at least from that which was available) would be of little use in either an overlay scheme or a direct classification scheme that would use anything other than relative statistics. This is indicated from the non-absolute transfer function of the precision processing equipment.

11.0 Conclusions and Recommendations

The conclusions for each of the nine projects composing this study are included at the end of each of the sections. These conclusions will be summarized here and tied together with an overall view as to the outcome of the study and recommendations for future work are included.

The overall conclusions from the crop identification and acreage estimation phase of the investigation are that the combination of ERTS-1 MSS data and machine processing of it can be used to obtain crop acreage information over large areas of the world. It has been shown in this study that it is possible to accurately identify major crop species from ERTS data and to convert the identification data to accurate estimates of the crop acreage using machine processing methods. The best performance is obtained when the data is collected at the right time in the crop's growth cycle, the fields are relatively large and uniform, and there are not too many crops to be identified. On the other hand, if there are several crops having similar characteristics or if the area to be classified is heterogeneous in its composition of crops and condition, ERTS data may not have sufficient spectral bands to enable accurate identification of individual species.

The overall quality of the ERTS-1 MSS data was judged to be high; however, the 80 meter instantaneous field of view is a limitation in areas having small fields and the four bands are a minimum for producing accurate classifications. It is recommended that additional wavelength bands in the middle and thermal infrared be considered since they would undoubtedly improve classification performance, particularly in those areas having more than two or three major crops to be identified. Cloud cover may be a limiting factor in some instances, but in an operational environment where data were being analyzed from over

large areas (rather than small pre-designated test sites) this might very well be a less serious problem. Still, in some agricultural situations more frequent collection would add to the value of the data; thus it is recommended that more frequent coverage be considered in future systems. Also, the analysis of ERTS data is handicapped by the six to eight week interval between data collection and receipt of the data tapes and imagery. This is a particularly serious problem for agricultural crops which change quite rapidly and may even be harvested before the data is received. In order to carry out the best analysis, ground observation data needs to be collected very near the time of ERTS data collection. However, analysts are reluctant to spend a lot of time and effort collecting ground truth until they know that cloud-free ERTS data was collected. They are, therefore, faced with the choice of collecting ground observation data which may never be used or trying to collect the necessary data after it may be too late. Neither alternative allows for optimum use of the ERTS data. It is thus recommended that "quick look" image products be made available much more rapidly than with the present system.

The overall conclusion from the soil association mapping project was that strong relationships exist between ERTS imagery and conventionally mapped generalized soil boundaries for all three test counties. Results indicate that computer analysis of MSS imagery provided better discrimination among soils than single band imagery or false color enhancement using multiple bands. A major limitation of the computer analysis was selection of training samples which were representative of the soils over a large area (501 square miles). Computer analysis of MSS data was more flexible than the photographic approach in several respects: (1) It facilitated analysis over smaller areas in more detail. (2) It was possible to select a data set from a small area (such as a county) rather than using an entire frame

of data. (3) The use of three spectral bands and simulation of color infrared photography have been shown to enhance differences among soils to a greater extent than single band photographic techniques. Use of color IR simulation techniques are subject to many of the same limitations which are inherent with conventional color and color IR aerial photography. These shortcomings of aerial photography in remote sensing are well known. Among other considerations, only linear combinations of wavelengths were possible in this study using the false color enhancement of ERTS MSS data and the interpretations which can be made are still largely subject to tonal differences rather than actual measured differences in multispectral reflectance.

Preliminary studies which were conducted in the late 1960's as remote sensing was becoming more involved in soil mapping pointed out that soil moisture, soil surface roughness, and other surface conditions not directly related to the mapping of soils had some effect on soil spectral characteristics. In this study no information was obtained as to the surface condition at the time the spectral data were gathered, other than to assure that the surfaces were nonvegetated. The results obtained from this study are particularly encouraging when it is considered that while soil surface conditions were confounded with soil properties of interest in mapping, it was still possible to separate soils into meaningful classes over a large area.

The conclusions of the Urban Land Use Analysis project suggest that computer analysis of ERTS MSS data may be a valuable tool for the urban-regional planner. Although only gross land use inventories may be made, because of the satellite's resolution, timely updating of a metropolitan area's data bank would be invaluable. Detection of land use change by the satellite would indicate where detailed studies (either by aerial photography or direct field investigation) ought to be pursued. Such detection

would have been possible before costly air photo coverage or ground observations of the entire area were made. At the state/regional or national levels, machine processing of ERTS data may be totally adequate. State officials are not disinterested in local land use problems, but they are concerned more with broad trends within a large area. The achieved classification accuracies of 87 to 92 per-cent may be adequate for their purposes.

The results of the Water Resources project indicate that ERTS-1 multispectral data and computer-aided classification techniques can be utilized to detect and map different spectral classes of surface-water which may correspond to different levels of turbidity. However, it is clear that more work in conjunction with collection of more accurate and reliable field observations are needed in this area of research. Nevertheless, from previously reported work on turbidity with ERTS-1 and EXOTECH field spectroradiometer data and from existing processing and analysis techniques at LARS, such as the "Layered Classifier", it seems feasible to be able to map and make quantitative determinations of water turbidity levels. The procedure recommended for the quantitative determination of the amount of suspended solids present in lakes and reservoirs is to utilize a layered classification scheme in which the first step would be to use all four spectral bands of ERTS for the separation of water from every other cover type through a maximum likelihood classification. Then, the second step would consist of a level slicing technique applied to only one spectral band, such as band 5 ($0.6 - 0.7\mu\text{m}$) which has been shown to have a linear response as a function of turbidity levels. In the case of the aircraft data analysis, the results indicate that under certain conditions the sun-scanner-look angle effect is so pronounced that the data is useless for surface-water studies because any spectral

characteristic due to either depth, turbidity or any other water quality parameter would be completely masked by the strong specular reflections from the water surface. Therefore, careful aircraft mission planning is needed in order to avoid the sun-scanner-look angle effects. Finally, it was shown that because of the coarse spatial resolution of the ERTS-1 sensor system, there is a consistent under-estimation of the surface area of water bodies. However, two correction methods were developed and successfully tested. One considers the water-edge spectral class, and the other takes into account the fact that the magnitude of under-estimation is a definite function of the size of the water bodies. The resulting corrected water acreage estimations using ERTS-1 MSS data together with computer-aided processing techniques, have been shown to be statistically correlated to the standard USGS data. Thus, we can conclude that it is possible to accurately estimate the size of lakes and reservoirs from ERTS-1 data, provided an appropriate correction function is applied.

The conclusion of the Earth Surface Features Identification is that computer analysis of ERTS data would make a valuable source of earth surface feature data for regional resource planning. An overall identification accuracy of 85% was achieved in this experiment. It is also found that the overall identification accuracy will also become greater with use of temporal classification techniques. Although this investigation was limited to comparison of only forest cover data, other earth surface resources could be identified and analyzed by the same process. The spatial accuracy results are encouraging and the future holds that with the achievement of a more accurate automatic ERTS data entry a semi-automatic system of data bank development for utilization in the land use planning process can result.

The most conclusive finding in Analysis Technique Development project has been that the multispectral data analysis techniques heretofore developed for aircraft data and, to a much lesser extent, digitized space photography can be effectively applied, with appropriate modifications, to multispectral scanner data from satellites. Most important of the modifications has been upgrading of the cluster analysis capability, which is used in both supervised and unsupervised analysis modes. A model has been developed for adaptive classification where large geographical areas are involved. However, the results of attempts to evaluate this model have been inconclusive, largely due to the unavailability of suitable test data. A method for "recursive image partitioning" has been implemented which utilizes scene context to decompose the scene into self-defined "objects". This technique appears to be of greatest potential use for reducing analysis results storage, although other applications related to unsupervised image analysis can also be envisioned. The application of layered decision logic has been demonstrated to be a potentially powerful tool for the design of future pattern classifier systems. This approach offers not only improved efficiency (speed), but also makes optimal use of available information to maximize classification accuracy. Under this contract, promising research results have been obtained and rudimentary software developed. This is a top priority area for further work.

The Reformatting and Temporal Overlay task was highly successful in providing a flow of reformatted and preprocessed data for CCT users throughout the study. Data quality assurance, temporal registration of multiple frames over the same area, geometric correction and special frame operations were performed

on user request for the various projects. All eight other projects utilized the services of the reformatting and overlay task area during the study. The conclusion drawn is that a refined digital data handling system is key to successful employment of CCT data and geometric correction and data correction are highly desirable preprocessing operations. Temporal overlay is required where increased classification accuracy can be achieved from the temporal dimension and for change analysis.

The Atmospheric Modeling Project developed a model for atmospheric effects and it was applied to ERTS data from the Northern Illinois area for August 1972. Corrections were developed for the four ERTS MSS bands and the data preprocessed with these corrections. Time did not permit classification accuracy comparison with the correction. This work is continuing under another sponsor.

The Comparison of System and Scene Corrected MSS Data project revealed serious problems in the radiometric and geometric quality of the scene corrected data. Thus this project did not achieve its desired goal of overlaying system corrected on scene corrected data. The project was terminated early in the study after the initial analysis was made and the resources applied to the other projects.

12. Acknowledgements

The research performed on this contract was guided by the nine project leaders and numerous LARS Staff employees, graduate students and undergraduates. A complete list is not included; however, particular thanks is given to the following contributors:

Project Management

Terry L. Phillips
Paul E. Anuta

Technical Monitor and Reporting Editor

Paul E. Anuta

Project Leaders

Marvin E. Bauer - Crop Identification
Jan E. Cipra - Soil Association Mapping
William Todd - Urban Land Use Analysis
Roger M. Hoffer - Water Resources Research
Richard A. Boots - Earth Surface Features Identification
Philip H. Swain - Analysis Technique Development
William R. Simmons - Reformatting
Paul E. Anuta - Temporal Overlay
Gerald M. Jurica - Atmospheric Modeling
Leroy F. Silva - Comparison of System and Scene-Corrected Data

Clerical and Report Production

Carl C. Westfall
Thomas C. Builta

Supporting Staff and Students

Luis Bartolucci
Jerry Leslie

Eric Yam
William Shelley
David Freeman
Jim Kast
Keith Philipp

13.0 Publications Resulting from the Study

A list of the special report, LARS Information Notes, Conference Papers, and journal papers generated as a result of the ERTS Wabash Valley Contract is presented here. Excluded are all Type I and II reports.

1. Landgrebe, D. A., Hoffer, R. M., Staff, "An Early Analysis of ERTS-1 Data", LARS Information Note 092972, September 29, 1972. Also, presented at 1st ERTS Program Review, September 1972.
2. Baumgardner, M. E., Todd, W. J., "Land-Use Classification of Marion County, Indiana by Spectral Analysis of Digitized Satellite Data", Presented at the LARS/Purdue Conference on Machine Processing of Remotely Sensed Data, October 16 - 18, 1973, Proceedings pages 2a - 23, 2a - 32.
3. Cipra, J. E., "Mapping Soil Associations Using ERTS MSS Data", *ibid*, pp. 3a - 1, 3a - 10.
4. Robertson, T. V., "Extraction and Classification of Objects in Multispectral Images", *ibid*, pp. 3b - 27, 3b - 34.
5. Cipra, J. E., Franzmeier, D. P., Steinhardt, "Mapping of Soil Characteristics by Computer Analysis of ERTS MSS Imagery", Paper presented at American Society of Agronomy Annual Meeting, Las Vegas, Nevada, November 1973.
6. Steinhardt, G. C., Franzmeier, D. P., Cipra, J. E., "Comparison of Indiana Soil Associations and Earth Resources Technology Satellite Imagery", Presented at same Conference as 5, no proceedings.
7. Anuta, P. E., "Geometric Correction of ERTS-1 Digital Multispectral Scanner Data", Information Note 103073. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

8. Bauer, M. E., Cipra, J. E., "Identification of Agricultural Crops by Computer Processing of ERTS MSS Data", 2nd NASA ERTS Applications Investigator Program Review, NASA Goddard Space Flight Center, Greenbelt, Maryland, March 5 - 7, 1973, Volume I, Section A, pp. 205 - 212.
9. Todd, W. J., Mausel, P., Wenner, K. A., "Preparation of Urban Land Use Inventories by Machine Processing of ERTS MSS Data", *ibid*, Volume I, Section B, pp. 1031 - 1039.
10. Landgrebe, D. A., Staff, "An Evaluation of Machine Processing Techniques of ERTS-1 Data User Applications", LARS Information Note 121373, December 1973; Presented at 3rd NASA ERTS Applications Investigator Program Review, December 10, 1973, Washington, D. C.

14. Image Descriptor Forms

ERTS IMAGE DESCRIPTOR FORM

(See Instructions on Back)

DATE January 30, 1974

PRINCIPAL INVESTIGATOR David A. Landgrebe

GSFC UN127

ORGANIZATION Laboratory for Applications of Remote Sensing
Purdue University

NDPF USE ONLY

D _____
 N _____
 ID _____

| PRODUCT ID (INCLUDE BAND AND PRODUCT) | FREQUENTLY USED DESCRIPTORS* | | | DESCRIPTORS |
|--|------------------------------|------|----------|---------------|
| | | | | |
| <u>Idaho</u> | | | | |
| 1034-17473 | Cropland | | | |
| 1035-17525 | Cropland | | | |
| <u>Illinois</u> | | | | |
| 1017-16093 | Cropland | Corn | Soybeans | |
| 1017-16100 | Cropland | Corn | Soybeans | |
| 1053-16095 | Cropland | Corn | Soybeans | |
| 1071-16095 | Cropland | Corn | Soybeans | |
| <u>Indiana</u> | | | | |
| 1394-16035 | Cropland | Corn | Soybeans | Lake Michigan |
| 1394-16042 | Cropland | Corn | Soybeans | |

*FOR DESCRIPTORS WHICH WILL OCCUR FREQUENTLY, WRITE THE DESCRIPTOR TERMS IN THESE COLUMN HEADING SPACES NOW AND USE A CHECK (✓) MARK IN THE APPROPRIATE PRODUCT ID LINES. (FOR OTHER DESCRIPTORS, WRITE THE TERM UNDER THE DESCRIPTORS COLUMN).

MAIL TO ERTS USER SERVICES
 CODE 563
 BLDG 23 ROOM E413
 NASA GSFC
 GREENBELT, MD. 20771
 301-982-5406

ERTS IMAGE DESCRIPTOR FORM

(See Instructions on Back)

DATE May 9, 1973

PRINCIPAL INVESTIGATOR David A. Landgrebe

GSFC UN127

ORGANIZATION Laboratory for Applications of Remote Sensing
Purdue University

NDPF USE ONLY

D _____
N _____
ID _____

| PRODUCT ID (INCLUDE BAND AND PRODUCT) | FREQUENTLY USED DESCRIPTORS* | | | DESCRIPTORS |
|--|------------------------------|---------------|-------|-----------------------|
| | Rural Area | Urban Area | Water | |
| 106915585MX ↓ | | | ✓ | Pond |
| | | | ✓ | Reservoirs |
| | | | ✓ | River |
| | | | ✓ | Stream |
| | | ✓ | | Industrial area |
| | | ✓ | | Commercial area |
| | | ✓ | | Older housing |
| | | ✓ | | Suburban area |
| | | ✓ | | Highways |
| | | ✓ | | Parks |
| | | ✓ | | Golf courses |
| | | ✓ | | Wooded suburban areas |
| | ✓ | | | Cropland |
| | ✓ | | | Wooded areas |
| | ✓ | ✓ | | Cumulus clouds |
| | ✓ | ✓ | | Cloud shadows |
| | | ✓ | | Airport |

*FOR DESCRIPTORS WHICH WILL OCCUR FREQUENTLY, WRITE THE DESCRIPTOR TERMS IN THESE COLUMN HEADING SPACES NOW AND USE A CHECK (✓) MARK IN THE APPROPRIATE PRODUCT ID LINES. (FOR OTHER DESCRIPTORS, WRITE THE TERM UNDER THE DESCRIPTORS COLUMN).

MAIL TO ERTS USER SERVICES
CODE 563
BLDG 23 ROOM E413
NASA GSFC
GREENBELT, MD. 20771
301-982-5406

ERTS IMAGE DESCRIPTOR FORM

(See Instructions on Back)

DATE May 7, 1973PRINCIPAL INVESTIGATOR D. A. LandgrebeGSFC UN 127ORGANIZATION Laboratory for Applications of Remote Sensing
Purdue University

NDPF USE ONLY

D _____

N _____

ID _____

| PRODUCT ID (INCLUDE BAND AND PRODUCT) | FREQUENTLY USED DESCRIPTORS* | | | DESCRIPTORS |
|--|------------------------------|--------|---------|-------------|
| | Cropland | Cotton | Soybean | |
| 1034-16052 | X | X | X | |
| 1034-16055 | X | X | X | |
| 1052-16052 | X | X | X | |
| 1052-16055 | X | X | X | |
| 1070-16052 | X | X | X | |
| 1070-16055 | X | X | X | |

*FOR DESCRIPTORS WHICH WILL OCCUR FREQUENTLY, WRITE THE DESCRIPTOR TERMS IN THESE COLUMN HEADING SPACES NOW AND USE A CHECK (✓) MARK IN THE APPROPRIATE PRODUCT ID LINES. (FOR OTHER DESCRIPTORS, WRITE THE TERM UNDER THE DESCRIPTORS COLUMN).

MAIL TO ERTS USER SERVICES
CODE 563
BLDG 23 ROOM E413
NASA GSFC
GREENBELT, MD. 20771
301-982-5406